



```

MEDLINE-93170315; PubMed=8094669;
Das A.F., Arnberg A.C., Malingre H., Moerer P., Charles R.,
Moorman A.F.M., Lamers W.H.;
"Isolation and characterization of the rat gene encoding glutamate
dehydrogenase.";
Eur. J. Biochem. 211:795-803(1993).
-1- CATALYTIC ACTIVITY: L-GLUTAMATE + H(2)O + NAD(P)(+) -
2-OXOGLUTARATE + NH(3) + NAD(P)H.
-1- SUBUNIT: HOMOHEXAMER.
-1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
-1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X14223; CAA32441.1; -
EMBL; X14044; CAA32202.1; -
EMBL; X64365; CAA45717.1; -
PIR; S03707; S03707.
HSSP; P80319; IGTM.
InterPro; IPR001625; -
Pfam; PF00208; GLFV_dehydrog; 1.
PRINTS; PR00082; GLFVHDRGNASE.
PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
Oxidoreductase; NADP; Mitochondrion; Transit peptide.
TRANSIT 1 53 MITOCHONDRION
CHAIN 54 58 SUBSTRATE.
BINDING 84 84
ACT_SITE 183 183
CONFLICT 56 57
SEQUENCE 558 AA; 61428 MW; 3E8B8B5490C4C14D CRC64;
GP -> EA (IN REF. 2).

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 2; Mismatches 7; Indels 4; Gaps 2;

Qy 4 GLYQGVPR-EPGTE--ARRHYDEG 25
|.:|:|:|:|:|:|:|:|:|
Db 31 GWARGQPSAVPQPGGLTPVARRHYSEG 56

RESULT 3
INVD_ZYMMO
ID INVD_ZYMMO STANDARD; PRT; 413 AA.
AC Q60115; Q60117; Q60125;
AT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR SUCRASE (EC 3.2.1.26) (BETA-FRUCTOFURANOSIDASE)
DE (INVERTASE) (SACCHARASE) (PROTEIN B46).
DE SACC OR INV OR SUCE3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
OC NCBI_TaxID=542;
OX [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=ATCC 10988 / ZML;
RC MEDLINE=94368848; PubMed=8086457;
RX Song K.B., Lee S.K., Joo H.K., Rhee S.-K.;
RA "Nucleotide and derived amino acid sequences of an extracellular
RT sucrose gene (invB) of Zymomonas mobilis ZM1 (ATCC10988).";
RL Biochim. Biophys. Acta 1219:163-166(1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=IFO 13756;
RC MEDLINE=95218269; PubMed=7766026;
RX

```





Query Match	32.2%;	Score 47;	DB 1;	Length 457;
Best Local Similarity	42.9%;	Pred No. 25;		
Matches. 12;	Conservative	4;	Mismatches	6; Gaps 1;
QY	3	EGLYQGVV-----RAEPGTEARRHYDE	24	
		:	:	:
Db	204	EGLSQGLVDREKGLSAEPGWQAKREEE	231	
RESULT 5				
DHE3_HUMAN				
ID	DHE3_HUMAN	STANDARD;	PRT;	558 AA.
AC	P00367;			
DT	21-JUL-1986	(Rel. 01,	Created)	
DT	01-JAN-1990	(Rel. 13,	Last sequence update)	









```
FT VARIANT 42 P -> S.
FT VARIANT 119 H -> R.
FT VARIANT 246 R -> S.
FT VARIANT 256 P -> L.
FT VARIANT 261 P -> A.
SQ SEQUENCE 531 AA; 57545 MW; 737339803F75AB19 CRC64;

Query Match
Best Local Similarity 31.5%; Score 46; DB 1; Length 531;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17
   | | | | |
Db 314 YEGAPNAEPTTE 325

RESULT 10
DHE4_HUMAN
ID DHE4_HUMAN STANDARD; PRT; 558 AA.
AC P49448;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUTAMATE DEHYDROGENASE 2 PRECURSOR (EC 1.4.1.3) (GDH).
GN GLUD2 OR GLUDP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=942666921; PubMed=8207021;
RA Shashidharan P., Michaelidis T.M., Robakis N.K., Kresovali A.,
RA Papamatheakis J., Plaitakis A.;
RT "Novel human glutamate dehydrogenase expressed in neural and
RT testicular tissues and encoded by an X-linked intronless gene.";
RL J. Biol. Chem. 269:16971-16976(1994).
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE + H(2)O + NAD(P)(+) -
CC 2- OXOGLUTARATE + NH(3) + NAD(P)H.
CC -1- SUBUNIT: HOMOHXAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HUMAN RETINA, TESTIS, AND, AT A
CC LOWER LEVEL, BRAIN.
CC -1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X66310; CAA46995.1; -
DR EMBL; U08997; AAA20969.1; -
DR HSSP; P80319; 1GTM.
DR MIM; 300144; -
DR InterPro; IPR001625; -
DR Pfam; PF00208; GLFV_dehydrog; 1.
DR PRINTS; PR00082; GLFDHGRGNASE.
DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Mitochondrion; Transit peptide;
KW Multigene family.
FT TRANSIT 1 53 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 54 558 GLUTAMATE DEHYDROGENASE 2.
FT BINDING 84 84 SUBSTRATE (BY SIMILARITY).
FT ACT_SITE 183 183 BY SIMILARITY.
SQ SEQUENCE 558 AA; 61433 MW; F4E5B072C4EF32EA CRC64;

Query Match
31.5%; Score 46; DB 1; Length 558;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Best Local Similarity 42.3%; Pred. No. 42;
Matches 11; Conservative 3; Mismatches 10; Indels 2; Gaps 1;

QY 1 VEGLYQGVPRAPG--TEARRHYDE 24
   : | | | : | | | |
Db 30 LGRGRGQPAASQPLALARRHYSE 55

RESULT 11
YBHF_ECOLI
ID YBHF_ECOLI STANDARD; PRT; 578 AA.
AC P75776;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBHF.
GN YBHF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE000181; AAC73881.1; ALT_INIT.
DR EcoGene; EG13314; ybhF.
DR InterPro; IPR001617; -
DR Pfam; PF00005; ABC_tran; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Hypothetical protein; ATP-binding; Transport.
FT NP_BIND 40 47 ATP (POTENTIAL).
FT NP_BIND 362 369 ATP (POTENTIAL).
SQ SEQUENCE 578 AA; 63132 MW; DB3B3FA213490F3C CRC64;

Query Match
31.5%; Score 46; DB 1; Length 578;
Best Local Similarity 52.4%; Pred. No. 44;
Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEARRHY 22
   | | | | | | |
Db 213 GELLYQGEKALTQTMAGRSF 233

RESULT 12
MENI_HUMAN
ID MENI_HUMAN STANDARD; PRT; 615 AA.
AC O00255; O00632;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MENIN.
GN MENI OR SCG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANTS FMEN1.  
RC TISSUE=Leukocyte;  
RX MEDLINE=97258940; PubMed=9103196;  
RA Chandrasekharappa S.C., Guru S.C., Manickam P., Olufemi S.-E.,  
RA Collins F.S., Emmert-Buck M.R., Debelenko L.V., Zhuang Z.,  
RA Lubensky I.A., Liotta L.A., Crabtree J.S., Wang Y., Roe B.A.,  
RA Weisemann J., Boguski M.S., Agarwal S.K., Kester M.B., Kim Y.S.,  
RA Heppner C., Dong Q., Spiegel A.M., Burns A.L., Marx S.J.;  
RT "Positional cloning of the gene for multiple endocrine neoplasia-type  
1.";  
RL Science 276:404-407(1997).  
RN [2]  
RP VARIANTS FMEN1 AND SPORADIC MEN1.  
RX MEDLINE=97358593; PubMed=9215689;  
RA Agarwal S.K., Kester M.B., Debelenko L.V., Heppner C.,  
RA Emmert-Buck M.R., Skarulis M.C., Doppman J.L., Kim Y.S.,  
RA Lubensky I.A., Zhuang Z., Green J.S., Guru S.C., Manickam P.,  
RA Olufemi S.E., Liotta L.A., Chandrasekharappa S.C., Collins F.S.,  
RA Spiegel A.M., Burns A.L., Marx S.J.;  
RT "Germline mutations of the MEN1 gene in familial multiple endocrine  
neoplasia type 1 and related states.";  
RL Hum. Mol. Genet. 6:1169-1175(1997).  
RN [3]  
RP VARIANT PARATHYROID ADENOMA LYS-26.  
RX MEDLINE=97385243; PubMed=9241276;  
RA Heppner C., Kester M.B., Agarwal S.K., Debelenko L.V.,  
RA Emmert-Buck M.R., Guru S.C., Manickam P., Olufemi S.-E.,  
RA Skarulis M.C., Doppman J.L., Alexander R.H., Kim Y.S., Saggar S.K.,  
RA Lubensky I.A., Zhuang Z., Liotta L.A., Chandrasekharappa S.C.,  
RA Collins F.S., Spiegel A.M., Burns A.L., Marx S.J.;  
RT "Somatic mutation of the MEN1 gene in parathyroid tumours.";  
RL Nat. Genet. 16:375-378(1997).  
RN [4]  
RP VARIANTS FMEN1 D-42; P-165; D-169; S-188 AND E-289.  
RX MEDLINE=98130524; PubMed=9463336;  
RA Bassett J.H.D., Forbes S.A., Pannett A.A.J., Lloyd S.E.,  
RA Christie P.T., Wooding C., Harding B., Besser G.M., Edwards C.R.,  
RA Monson J.P., Sampson J., Wass J.A.H., Wheeler M.H., Thakker R.V.;  
RT "Characterization of mutations in patients with multiple endocrine  
neoplasia type 1.";  
RL Am. J. Hum. Genet. 62:232-244(1998).  
RN [5]  
RP VARIANTS FMEN1.  
RX MEDLINE=98349969; PubMed=96833585;  
RA Giraud S., Zhang C.X., Serova-Sinilnikova O., Wautot V., Salandre J.,  
RA Buisson N., Waterlot C., Bautres C., Porchet N., Aubert J.-P., Emy P.,  
RA Cadiot G., Delemer B., Chabre O., Niccoli P., Lepirat F., Duron F.,  
RA Emperauger B., Cougard P., Goudet P., Sarfati E., Riou J.-P.,  
RA Guichard S., Rodier M., Meyrier A., Caron P., Vantighem M.-C.,  
RA Assayag M., Peix J.-L., Pugeat M., Rohmer V., Vallotton M., Lenoir G.,  
RA Gaudray P., Proye C., Conte-Devolx B., Chanson P., Shugart Y.Y.,  
RA Goldgar D., Murat A., Calender A.;  
RT "Germ-line mutation analysis in patients with multiple endocrine  
neoplasia type 1 and related disorders.";  
RL Am. J. Hum. Genet. 63:455-467(1998).  
RN [6]  
RP VARIANT FIHP LYS-260.  
RX MEDLINE=99011276; PubMed=9792884;  
RA Teh B.T., Esapa C.T., Houlston R., Grandell U., Farnebo F.,  
RA Nordenskjoeld M., Pearce C.J., Carmichael D., Larsson C., Harris P.E.;  
RT "A family with isolated hyperparathyroidism segregating a missense  
MEN1 mutation and showing loss of the wild-type alleles in the  
parathyroid tumors.";  
RL Am. J. Hum. Genet. 63:1544-1549(1998).  
RN [7]  
RP VARIANT FIHP GLU-189.  
RX MEDLINE=99057176; PubMed=9843042;  
RA Fujimori M., Shirahama S., Sakurai A., Hashizume K., Hama Y., Ito K.,  
RA Shingu K., Kobayashi S., Amano J., Fukushima Y.;  
RT "Novel V184E MEN1 germline mutation in a Japanese kindred with

RT familial hyperparathyroidism.";  
RL Am. J. Med. Genet. 80:221-222(1998).  
RN [8]  
RP VARIANTS FMEN1.  
RX MEDLINE=98334342; PubMed=9671267;  
RA Agarwal S.K., Debelenko L.V., Kester M.B., Guru S.C., Manickam P.,  
RA Olufemi S.-E., Skarulis M.C., Heppner C., Crabtree J.S.,  
RA Lubensky I.A., Zhuang Z., Kim Y.S., Chandrasekharappa S.C.,  
RA Collins F.S., Liotta L.A., Spiegel A.M., Burns A.L., Emmert-Buck M.R.,  
RA Marx S.J.;  
RT "Analysis of recurrent germline mutations in the MEN1 gene encountered  
in apparently unrelated families.";  
RL Hum. Mutat. 12:75-82(1998).  
RN [9]  
RP VARIANT FMEN1 ILE-135 AND LYS-364.  
RX MEDLINE=98410971; PubMed=9740255;  
RA Boeni R., Vortmeyer A.O., Pack S., Park W.-S., Burg G., Hofbauer G.,  
RA Darling T., Liotta L., Zhuang Z.;  
RT "Somatic mutations of the MEN1 tumor suppressor gene detected in  
sporadic angiofibromas.";  
RL J. Invest. Dermatol. 111:539-540(1998).  
RN [10]  
RP VARIANTS FMEN1 LYS-119 DEL AND GLN-171--LEU-173 DEL.  
RX MEDLINE=98419173; PubMed=9747036;  
RA Sakurai A., Shirahama S., Fujimori M., Katai M., Itakura Y.,  
RA Kobayashi S., Amano J., Fukushima Y., Hashizume K.;  
RT "Novel MEN1 gene mutations in familial multiple endocrine neoplasia  
type 1.";  
RL J. Hum. Genet. 43:199-201(1998).  
RN [11]  
RP VARIANT FMEN1 GLY-45.  
RX MEDLINE=99048878; PubMed=9832038;  
RA Sato M., Matsubara S., Miyauchi A., Ohye H., Imachi H., Murao K.,  
RA Takahara J.;  
RT "Identification of five novel germline mutations of the MEN1 gene in  
Japanese multiple endocrine neoplasia type 1 (MEN1) families.";  
RL J. Med. Genet. 35:915-919(1998).  
RN [12]  
RP VARIANTS FMEN1 TRP-39; TYR-177; ASP-184; PRO-269 AND PRO-272.  
RX MEDLINE=99103464; PubMed=9888389;  
RA Poncin J., Abs R., Velkeniers B., Bonduelle M., Abramowicz M.,  
RA Legros J.-J., Verloes A., Meurisse M., van Gaal L., Verellen C.,  
RA Koulischer L., Beckers A.;  
RT "Mutation analysis of the MEN1 gene in Belgian patients with multiple  
endocrine neoplasia type 1 and related diseases.";  
RL Hum. Mutat. 13:54-60(1999).  
RN [13]  
RP VARIANTS MEN1 ASP-161 AND ARG-246.  
RX MEDLINE=99188881; PubMed=10090472;  
RA Mutch M.G., Dilley W.G., Sanjurjo F., Debenedetti M.K., Doherty G.M.,  
RA Wellis S.A. Jr., Goodfellow P.J., Lairmore T.C.;  
RT "Germline mutations in the multiple endocrine neoplasia type 1 gene:  
evidence for frequent splicing defects.";  
RL Hum. Mutat. 13:175-185(1999).  
CC -!- FUNCTION: NOT KNOWN.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.  
CC -!- DISEASE: DEFECTS IN MEN1 IS THE CAUSE OF FAMILIAL MULTIPLE  
CC -!- DISEASE: DEFECTS IN MEN1 IS THE CAUSE OF FAMILIAL ISOLATED  
CC DOMINANT DISORDER CHARACTERIZED BY TUMORS OF THE PARATHYROID  
CC GLANDS, GASTRO-INTESTINAL ENDOCRINE TISSUE, THE ANTERIOR PITUITARY  
CC AND OTHER TISSUES. CUTANEOUS LESIONS AND NERVOUS-TISSUE TUMORS  
CC CAN EXIST. PROGNOSIS IN FMEN1 PATIENTS IS RELATED TO HORMONAL  
CC HYPERSECRETION BY TUMORS, SUCH AS HYPERGASTRINEMIA CAUSING SEVERE  
CC PEPTIC ULCER DISEASE (ZOLLINGER-ELLISON SYNDROME, ZES), PRIMARY  
CC HYPERPARATHYROIDISM, AND ACUTE FORMS OF HYPERINSULINEMIA.  
CC -!- DISEASE: DEFECTS IN MEN1 IS THE CAUSE OF FAMILIAL ISOLATED  
CC HYPERPARATHYROIDISM (FIHP OR HRPT1). FIHP IS AN AUTOSOMAL DOMINANT  
CC DISORDER CHARACTERIZED BY HYPERCALCEMIA, ELEVATED PARATHYROID  
CC HORMONE (PTH) LEVELS, AND UNIGLANDULAR OR MULTIGLANDULAR  
CC PARATHYROID TUMORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; U93236; AAC51228.1; -  
CC EMBL; U93237; AAC51229.1; -  
CC EMBL; U93237; AAC51230.1; -  
CC MIM; 131100; -  
CC MIM; 145000; -  
CC Nuclear protein; Disease mutation; Alternative splicing; Polymorphism.  
CC VARSPLIC 149 153  
CC MISSING (IN SHORT ISOFORM).  
CC P -> L (IN FMEN1).  
CC /FTid=VAR\_005425.  
CC L -> R (IN FMEN1).  
CC /FTid=VAR\_005426.  
CC E -> K (IN PARATHYROID ADENOMA).  
CC /FTid=VAR\_005427.  
CC L -> W (IN FMEN1).  
CC /FTid=VAR\_005428.  
CC G -> D (IN FMEN1).  
CC /FTid=VAR\_005429.  
CC E -> G (IN FMEN1).  
CC /FTid=VAR\_005430.  
CC MISSING (IN FMEN1).  
CC /FTid=VAR\_005431.  
CC K -> I (IN FMEN1).  
CC /FTid=VAR\_005434.  
CC H -> D (IN FMEN1).  
CC /FTid=VAR\_005432.  
CC H -> Y (IN FMEN1 AND SPORADIC MEN1).  
CC /FTid=VAR\_005433.  
CC F -> V (IN FMEN1).  
CC /FTid=VAR\_005436.  
CC G -> D (IN FMEN1).  
CC /FTid=VAR\_008017.  
CC A -> P (IN FMEN1).  
CC /FTid=VAR\_005437.  
CC A -> D (IN FMEN1).  
CC /FTid=VAR\_005438.  
CC MISSING (IN FMEN1).  
CC /FTid=VAR\_005439.  
CC R -> Q.  
CC

Query Match 31.5%; Score 46; DB 1; Length 615;  
Best Local Similarity 52.9%; Pred. No. 47;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GEGYQGVPRAPGTEA 18  
Db 512 GOGAVSGPPRKPPTGTA 528

RESULT 13  
ID P73\_CERAE STANDARD; PRT: 637 AA.  
AC Q9XSK8; Q9TSQ9;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE TUMOR PROTEIN P73 (P53-LIKE TRANSCRIPTION FACTOR) (P53-RELATED  
DE PROTEIN).  
DE TP73 OR P73.  
GN Cercopithecus aethiops (Green monkey) (Grivet).  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
OC Cercopithecinae; Cercopithecus.  
OX NCBI\_TaxID=9534;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;

RA Caput D.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
CC FUNCTION: PARTICIPATES IN THE APOPTOTIC RESPONSE TO DNA DAMAGE.  
CC WHEN OVERPRODUCED, ACTIVATES TRANSCRIPTION FROM P53-RESPONSIVE  
CC PROMOTERS AND INDUCES APOPTOSIS. MAY BE A TUMOR SUPPRESSOR  
CC PROTEIN (BY SIMILARITY).  
CC SUBUNIT: THE C-TERMINAL OLIGOMERIZATION DOMAIN BINDS TO THE ABL  
CC TYROSINE KINASE SH3 DOMAIN. ISOFORM BETA INTERACTS HOMOTYPICALLY  
CC AND WITH P53, WHEREAS ISOFORM ALPHA DOES NOT.  
CC SUBCELLULAR LOCATION: NUCLEAR.  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC DOMAIN: POSSESSES AN ACIDIC TRANSACTIVATION DOMAIN, A CENTRAL DNA  
CC BINDING DOMAIN AND A C-TERMINAL OLIGOMERIZATION DOMAIN THAT BINDS  
CC TO THE ABL TYROSINE KINASE SH3 DOMAIN.  
CC SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; Y11419; CAA72224.1; -  
CC EMBL; Y11419; CAA72225.1; -  
CC HSSP; P04637; LYCS.  
CC InterPro; IPR002117; -  
CC Pfam; PF00870; P53; 1.  
CC PRINTS; PR00386; P53SUPPRESSOR.  
CC Transcription regulation; Activator; DNA-binding; Anti-oncogene;  
CC Apoptosis; Nuclear protein; Phosphorylation; Alternative splicing.  
CC VARSPLIC 495 637  
CC SFLTGLGCPNCIEVFTSQGLQSIYHLONLTIEDLGALKIPE  
CC QYRMTIWRGLQDLKQGHYGAQAQQLLRSSNAAAISIGSG  
CC ELQQRVMEAVHFRVHTITIPNRGGPCAGPDEWAEFGDL  
CC PDCRKARKQPIKEEFTEAEIH -> RTWGP (IN ISOFORM  
CC BETA).  
CC SEQUENCE 637 AA; 69630 MW; 7CB200B919C9C70A CRC64;

Query Match 31.5%; Score 46; DB 1; Length 637;  
Best Local Similarity 56.2%; Pred. No. 48;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 9 VPRAEPTGEARRHYDE 24  
Db 337 VPALGPGVKRRHGDE 352

RESULT 14  
ID IE63\_HCMVA STANDARD; PRT: 744 AA.  
AC IE63\_HCMVA  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE TRANSCRIPTIONAL REGULATOR IE63 HOMOLOG (PROTEIN UL69).  
GN UL69.  
OS Human cytomegalovirus (strain AD169).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90269039; PubMed=2161319;  
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
RA Horsnell T., Hutchinson C.A., III, Kouzarides T., Martignetti J.A.,  
RA Freddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
RA "Analysis of the protein-coding content of the sequence of human  
RT Cytomegalovirus strain AD169."  
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).  
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,



CC HSV-2 UL54, EHV-1 5, VZV 4, EBV BMLF1, HCMV UL69, AND HVS-1 57.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X17403; CAA35384.1; -.  
DR PIR; S09832; QQBEA7.  
KW Transcription regulation.  
SQ SEQUENCE 744 AA; 82678 MW; D0566B2009EB5828 CRC64;

Query Match 31.5%; Score 46; DB 1; Length 744;  
Best Local Similarity 43.8%; Pred. NO. 56;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 10 PRAEPGTEARRHYDEG 25  
| | | : : : | | |  
Db 598 PPAQPPSQQQHYSEG 613

RESULT 15  
ILVE\_SALTY STANDARD; PRT; 308 AA.  
AC P15168;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE BRANCHED-CHAIN AMINO ACID AMINOTRANSFERASE (EC 2.6.1.42) (TRANSAMINASE  
DE B) (BCAT).  
GN ILVE.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=89352621; PubMed=2669973;  
RA Feild M.J., Nguyen D.C., Armstrong F.B.;  
RT "Amino acid sequence of Salmonella typhimurium branched-chain amino  
RT acid aminotransferase.";  
RL Biochemistry 28:5306-5310(1989).  
RN [2]

RP SEQUENCE OF 1-5.  
RA Randall R.R., Wallis M.H., Young G.J., Armstrong F.B.;  
RT "N-terminal sequence of branched-chain amino acid aminotransferase.";  
RL Fed. Proc. 38:325-325(1979).  
CC -!- CATALYTIC ACTIVITY: L-LEUCINE + 2-OXOGLUTARATE -> 4-METHYL-2-  
CC OXOPENTANOATE + L-GLUTAMATE (ALSO ACTS ON L-ISOLEUCINE AND  
CC L-VALINE).  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -!- PATHWAY: VALINE AND ISOLEUCINE BIOSYNTHESIS.  
CC -!- SUBUNIT: HOMOHXAMER.  
CC -!- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
DR PIR; A05077; A05077.  
DR PIR; A34082; A34082.  
DR HSSP; P00510; IA3G.  
DR StyGene; SG10181; ilve.  
DR InterPro; IPR001544; -.  
DR Pfam; PF01063; aminotran\_4; 1.  
DR PROSITE; PS00770; AA\_TRANSFER\_CLASS\_4; 1.  
KW Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;  
KW Pyridoxal phosphate.  
FT BINDING 159 159 PYRIDOXAL PHOSPHATE.  
SQ SEQUENCE 308 AA; 33921 MW; 8F826434F9588AAD CRC64;

Query Match 31.2%; Score 45.5; DB 1; Length 308;

Best Local Similarity 31.4%; Pred. NO. 27;  
Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;  
QY 3 EGLYQGV-----RAEP-----GTEARRH-YDEGV 26  
| | | : : : | | |  
Db 133 EALDQIDAMVSSWNRAAPTPTAAKAGGNYSLLVGCSEARRHGYQEGI 183

RESULT 16  
EST5\_DROMO STANDARD; PRT; 38 AA.  
ID EST5\_DROMO  
AC P10095;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE ESTERASE 5 (EC 3.1.1.1) (FRAGMENT).  
GN EST5 OR EST-5.  
OS Drosophila mojavensis (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7230;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=87100025; PubMed=3800957;  
RA Pen J., van Beeumen J., Beintema J.J.;  
RT "Structural comparison of two esterases from Drosophila mojavensis  
RT isolated by immunoaffinity chromatography.";  
RL Biochem. J. 238:691-699(1986).  
CC -!- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O -> AN ALCOHOL  
CC + A CARBOXYLIC ANION.  
CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.  
DR PIR; B29502; B29502.  
DR FlyBase; FBgn0012569; Dmoj\Est-5.  
DR InterPro; IPR002018; -.  
DR Pfam; PF00135; Coesterase; 1.  
DR PROSITE; PS00122; CARBOXYLESTERASE\_B\_1; PARTIAL.  
DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; PARTIAL.  
KW Hydrolase; Serine esterase.  
FT NON\_TER 38 38  
SQ SEQUENCE 38 AA; 4008 MW; 61A3C46CF084A465 CRC64;

Query Match 30.8%; Score 45; DB 1; Length 38;  
Best Local Similarity 64.3%; Pred. NO. 4;  
Matches 9; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 3 EGLY--QGVPRAP 14  
| | | : | | | | |  
Db 22 EGYEAEIGIPRAEP 35

RESULT 17  
FASA\_RAT STANDARD; PRT; 324 AA.  
ID FASA\_RAT  
AC Q63199;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE FASL RECEPTOR PRECURSOR (APOPTOSIS-MEDIATING SURFACE ANTIGEN FAS)  
DE (APO-1 ANTIGEN) (CD95).  
GN TNFRSF6 OR P1 OR FAS.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=94128114; PubMed=7507668;  
RA Kimura K., Yamamoto M., Wakatsuki T.;  
RT "A variant mRNA species encoding a truncated form of Fas antigen in  
RT the rat liver.";



RP SEQUENCE FROM N.A.  
RX MEDLINE=8087070; PubMed=3693370;  
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RT "Complete primary structure of the rat cartilage proteoglycan core  
protein deduced from cDNA clones.";  
RL J. Biol. Chem. 262:17757-17767(1987).  
RN [2]  
RP REVISION TO 698.  
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RL J. Biol. Chem. 263:10040-10040(1988).  
RN [3]  
RP SEQUENCE OF 1856-2124 FROM N.A.  
RX MEDLINE=86250698; PubMed=2424893;  
RA Doege K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;  
RT "Partial cDNA sequence encoding a globular domain at the C terminus  
of the rat cartilage proteoglycan.";  
RL J. Biol. Chem. 261:8108-8111(1986).  
CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR  
MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN  
IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO  
HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A  
REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.  
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY  
SIMILARITY).  
CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO  
TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,  
MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS  
CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS  
THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)  
AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2  
AND G3.  
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE  
CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 4 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M13518; AAA41836.1; -.  
DR EMBL; J03485; AAA21000.1; ALT\_SEQ.  
DR PIR; A23835; A23835.  
DR PIR; A28452; A28452.  
DR HSSP; P98066; 1TSG.  
DR InterPro; IPR000436; -.  
DR InterPro; IPR000495; -.  
DR InterPro; IPR000538; -.  
DR InterPro; IPR001304; -.  
DR Pfam; PF00193; xlink; 4.  
DR Pfam; PF00059; lectin\_c; 1.  
DR Pfam; PF00084; sushi; 1.  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PROSITE; PS01241; LINK; 4.  
DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; 1.  
DR PROSITE; PS50041; C\_TYPE\_LECTIN\_2; 1.  
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;  
KW Repeat; Immunoglobulin domain.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 2124 AGGREGAN CORE PROTEIN.  
FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 170 247 LINK 1.  
FT DOMAIN 268 349 LINK 2.  
FT DOMAIN 504 581 LINK 3.  
FT DOMAIN 602 683 LINK 4.  
FT DOMAIN 1910 2036 C-TYPE LECTIN.

FT DOMAIN 2040 2098 SUSHI.  
FT DOMAIN 48 140 G1-A.  
FT DOMAIN 152 247 G1-B.  
FT DOMAIN 253 349 G1-B'.  
FT DOMAIN 486 580 G2-B.  
FT DOMAIN 587 682 G2-B'.  
FT DOMAIN 685 798 KS.  
FT DOMAIN 801 1226 CS-1.  
FT DOMAIN 1227 1909 CS-2.  
FT DOMAIN 1910 2124 G3.  
FT DISULFID 51 133 BY SIMILARITY.  
FT DISULFID 175 246 BY SIMILARITY.  
FT DISULFID 199 220 BY SIMILARITY.  
FT DISULFID 273 348 BY SIMILARITY.  
FT DISULFID 297 318 BY SIMILARITY.  
FT DISULFID 509 580 BY SIMILARITY.  
FT DISULFID 533 554 BY SIMILARITY.  
FT DISULFID 607 682 BY SIMILARITY.  
FT DISULFID 631 652 BY SIMILARITY.  
FT DISULFID 1914 1925 BY SIMILARITY.  
FT DISULFID 1942 2034 BY SIMILARITY.  
FT DISULFID 2010 2026 BY SIMILARITY.  
FT DISULFID 2041 2084 BY SIMILARITY.  
FT DISULFID 2070 2097 BY SIMILARITY.  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1842 1842 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 2124 AA; 221117 MW; E30BBE61593A34B1 CRC64;  
  
Query Match 30.8%; Score 45; DB 1; Length 2124;  
Best Local Similarity 43.5%; Pred. NO. 2.2e+02;  
Matches 10; Conservative 1; Mismatches 12; Indels 0; Gaps 0;  
  
QY 4 GLYQGVPRAEPTGTEARRHYDEGV 26  
| :| | | | |  
Db 1100 GYVSGIPSGEDGTETSTSGVEGV 1122  
  
RESULT 20  
KAD\_PARDE  
ID KAD\_PARDE STANDARD; PRT; 217 AA.  
AC P10772; P77922;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).  
GN ADK.  
OS Paracoccus denitrificans.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Paracoccus.  
OX NCBI\_TaxID=266;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=71.11T;  
RX MEDLINE=99077027; PubMed=9862211;  
RA Perrier V., Burlacu-Miron S., Boussac A., Meier A., Gilles A.M.;  
RT "Metal chelating properties of adenylate kinase from Paracoccus  
denitrificans.";  
RL Protein Eng. 11:917-923(1998).  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=89153098; PubMed=2537726;  
RA Spurgin P., Tomasselli A.G., Schiltz E.;  
RT "The amino acid sequence of adenylate kinase from Paracoccus  
denitrificans and its relationship to mitochondrial and microbial  
adenylate kinases.";  
RL Eur. J. Biochem. 179:621-628(1989).  
CC -!- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR





FT DOMAIN 517 580 IG-LIKE C2-TYPE DOMAIN.  
FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 740 AA; 82378 MW; F312DC62C4B4A217 CRC64;  
  
Query Match 30.1%; Score 44; DB 1; Length 740;  
Best Local Similarity 63.6%; Pred. No. 1.1e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 12 AEPGTEARRHY 22  
Db 655 SDPNTTEARRHY 665  
  
RESULT 23  
RPA1\_HALHA STANDARD; PRT; 971 AA.  
AC P15350;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT A' (EC 2.7.7.6).  
GN RPOA1.  
OS Halobacterium halobium.  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
OX NCBI\_TaxID=2241;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=89199633; PubMed=2495365;  
RA Leffers H., Gropp F., Lottspeich F., Zillig W., Garrett R.A.;  
RT "Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaeobacterial extreme halophiles Halobacterium halobium and Halococcus morrhuae.";  
RL J. Mol. Biol. 206:1-17(1989).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.  
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE + RNA(N).  
CC -!- COFACTOR: ZINC.  
CC -!- SUBUNIT: THIS RNA POLYMERASE IS COMPOSED OF 5 LARGE SUBUNITS: A', A'', B', B'' AND EPSILON, PRESENT IN STOICHIOMETRIC AMOUNTS, AND 3 SMALL ONES, E, F AND G, PRESENT IN SUB-STOICHIOMETRIC AMOUNTS.  
CC -!- SIMILARITY: THE COMBINED A'+A'' SUBUNITS CORRESPOND TO THE A SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE EUBACTERIAL BETA' SUBUNIT.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X57144; CAA40426.1; -.  
PIR; S03574; S03574.  
DR InterPro; IPR000722; -.  
DR Pfam; PF00623; RNA\_pol\_A; 1.  
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc.  
FT MOD\_RES 1 1 BLOCKED.  
FT ZN\_FING 62 105 C4-TYPE (POTENTIAL).

SQ SEQUENCE 971 AA; 108683 MW; FA91E23E8DEA9A84 CRC64;  
  
Query Match 30.1%; Score 44; DB 1; Length 971;  
Best Local Similarity 53.3%; Pred. No. 1.4e+02;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
QY 11 RAEPGTEARRHYDEG 25  
Db 471 RVEPGWEVQRHLIDG 485  
I I I I I : I I : I  
  
RESULT 24  
VPG\_BYDVN STANDARD; PRT; 153 AA.  
ID VPG\_BYDVN  
AC P27579;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PUTATIVE GENOME-LINKED PROTEIN PRECURSOR (VPG) (17 KDA PROTEIN).  
OS Barley yellow dwarf virus (isolate NY-RPV) (BYDV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
OC Luteovirus.  
OX NCBI\_TaxID=12039;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91108372; PubMed=2273382;  
RA Vincent J.R., Ueng P.P., Lister R.M., Larkins B.A.;  
RT "Nucleotide sequences of coat protein genes for three isolates of barley yellow dwarf virus and their relationships to other luteovirus coat protein sequences.";  
RT J. Gen. Virol. 71:2791-2799(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92013950; PubMed=1840612;  
RA Larkins B.A., Lister R.M., Vincent J.R.;  
RT "Nucleotide sequence analysis and genomic organization of the NY-RPV isolate of barley yellow dwarf virus.";  
RT J. Gen. Virol. 72:2347-2355(1991).  
CC -!- FUNCTION: THIS PROTEIN IS POSSIBLY IS A VPG-PRECURSOR FROM WHICH, AT THE ONSET OF THE RNA SYNTHESIS, THE VPG MOLECULE IS RELEASED, AS IT HAS BEEN SUGGESTED FOR COWPEA MOSAIC VIRUS.  
CC -!- SIMILARITY: BELONGS TO THE LUTEOVIRUSES VPG PROTEIN FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D10206; BAA01056.1; -.  
DR EMBL; X17259; CAA35161.1; -.  
DR EMBL; L25299; AAA42870.1; -.  
DR InterPro; IPR001964; -.  
DR Pfam; PF01659; Luteo\_Vpg; 1.  
DR PRINTS; PR00912; LVIRUSORF5.  
SQ SEQUENCE 153 AA; 17209 MW; 3839E5C502C51602 CRC64;  
  
Query Match 29.8%; Score 43.5; DB 1; Length 153;  
Best Local Similarity 41.7%; Pred. No. 26;  
Matches 10; Conservative 3; Mismatches 4; Indels 7; Gaps 1;  
  
QY 1 VGEGLYQ-----GVPRAEPGTE 17  
: I I I I I : I I I I I  
Db 13 LGEGLLQERSQWLMSLPTAQPGAE 36  
  
RESULT 25  
ASGX\_PYRAB STANDARD; PRT; 305 AA.  
ID ASGX\_PYRAB

```

AC Q9V262:
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE PUTATIVE L-ASPARAGINASE (EC 3.5.1.1) (L-ASPARAGINE AMIDOHYDROLASE).
GN PAB0145.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
ON NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: L-ASPARAGINE + H(2)O -> L-ASPARTATE + NH(3).
CC -!- SIMILARITY: BELONGS TO THE ASPARAGINASE 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ248283; CAB49136.1; -
CC InterPro; IPR000246; -
CC HSSP; O47898; IAYV.
CC DR InterPro; IPR000246; -
CC DR Pfam; PF01112; Asparaginase_2; 1.
CC DR Hypothetical protein; Hydrolase.
CC SQ SEQUENCE 305 AA; 32591 MW; 5B2A1DE8F6FB2F7B CRC64;

Query Match 29.8%; Score 43.5; DB 1; Length 305;
Best Local Similarity 44.0%; Pred. No. 52;
Matches 11; Conservative 2; Mismatches 7; Indels 5; Gaps 1;

QY 1 VGEGLYQ-----GVPRAPGTEARR 20
DB 116 VEGAVKFAKIMGFPEYDPTTEARR 140
||||| : | : | | | | |
-!- VEGAVKFAKIMGFPEYDPTTEARR 140

RESULT 26
RECG_BORBU
ID RECG_BORBU STANDARD; PRT; 686 AA.
AC O51528;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE ATP-DEPENDENT DNA HELICASE RECG (EC 3.6.1.-).
GN RECG OR B0581.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
ON NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi.";
RL Nature 390:580-586(1997).
CC -!- FUNCTION: CRITICAL ROLE IN RECOMBINATION AND DNA REPAIR. HELP
CC -!- PROCESS HOLLIDAY JUNCTION INTERMEDIATES TO MATURE PRODUCTS BY

```

```

CC CATALYSING BRANCH MIGRATION. HAS A DNA UNWINDING ACTIVITY
CC CHARACTERISTIC OF A DNA HELICASE WITH A 3' TO 5' POLARITY. RECG
CC UNWIND BRANCHED DUPLEX DNA (Y-DNA) (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RECG SUBFAMILY OF HELICASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001159; AAC66942.1; -
CC TIGR; B00581; -
CC InterPro; IPR001410; -
CC InterPro; IPR001650; -
CC Pfam; PF00270; DEAD; 1.
CC DR Pfam; PF00271; Helicase_C; 1.
CC KW Helicase; DNA repair; ATP-binding; DNA recombination; DNA-binding.
CC FT NP_BIND 290 297 ATP (POTENTIAL).
CC FT SITE 391 394 DEOH BOX.
CC SQ SEQUENCE 686 AA; 79043 MW; 5D4B20F38E6BA50B CRC64;

Query Match 29.8%; Score 43.5; DB 1; Length 686;
Best Local Similarity 34.2%; Pred. No. 1.2e+02;
Matches 13; Conservative 3; Mismatches 7; Indels 15; Gaps 2;

QY 1 VEGEG-----LYQGVPRAPGTE-----ARRHYD 23
DB 292 VSGKTLVALLSGLPIEAGYQVAFMAPTDLLARQHYD 329
||||| : | | | | | | |
-!- VEGEG-----LYQGVPRAPGTE-----ARRHYD 23

RESULT 27
SYLM_NEUCR
ID SYLM_NEUCR STANDARD; PRT; 994 AA.
AC P15181;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE LEUCYL-TRNA SYNTHETASE, MITOCHONDRIAL PRECURSOR (EC 6.1.1.4)
DE (LEUCINE--TRNA LIGASE) (LEURS).
DE LEU-5.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
ON NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90097874; PubMed=2574823;
RA Chow C.M., Metznerberg R.L., Rajbhandary U.L.;
RA "Nuclear gene for mitochondrial leucyl-tRNA synthetase of Neurospora
crassa: isolation, sequence, chromosomal mapping, and evidence that
the leu-5 locus specifies structural information.";
RL Mol. Cell. Biol. 9:4631-4644(1989).
CC -!- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) = AMP +
CC PYROPHOSPHATE + L-LEUCYL-TRNA(LEU).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M30472; AAA33599.1; -
CC PIR; A33474; SYNCML.
CC InterPro; IPR001412; -
CC DR InterPro; IPR002300; -

```

DR InterPro; IPR002302; -.  
DR Pfam; PF00133; tRNA-synt.1; 1.  
DR PRINTS; PR00985; TRNASYNTHLEU.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Mitochondrion; Transit peptide.  
FT TRANSIT 1 52 MITOCHONDRION.  
FT CHAIN 53 994 LEUCYL-TRNA SYNTHETASE.  
FT SITE 93 103 "HIGH" REGION.  
FT SITE 697 701 "KMSKS" REGION.  
FT BINDING 700 700 ATP (BY SIMILARITY).  
SQ SEQUENCE 994 AA; 109853 MW; 2ADB26302A8465F9 CRC64;  
  
Query Match 29.8%; Score 43.5; DB 1; Length 994;  
Best Local Similarity 41.4%; Pred. No. 1.7e+02;  
Matches 12; Conservative 1; Mismatches 13; Indels 3; Gaps 1;  
  
QY 2 GEGLYQGVPRAPGTEA---RRHYDEGVR 27  
||| ||| : | ||| ||  
Db 393 GEGAVMGVPGHDLRDHAFWKEHHYDAPVR 421  
  
RESULT 28  
LRP1\_HUMAN STANDARD; PRT; 4544 AA.  
AC Q07954;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 1 PRECURSOR (LRP)  
DE (ALPHA-2-MACROGLOBULIN RECEPTOR) (A2MR) (APOLIPOPROTEIN E RECEPTOR)  
DE (APOER) (CD91).  
GN LRP1 OR A2MR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=89210795; PubMed=3266596;  
RA Herz J., Hamann U., Rogne S., Myklebost O., Gausepohl H.,  
RA Stanley K.K.;  
RT "Surface location and high affinity for calcium of a 500-kd liver  
RT membrane protein closely related to the LDL-receptor suggest a  
RT physiological role as lipoprotein receptor.";  
RL EMBO J. 7:4119-4127(1988).  
RN [2]  
RP PROCESSING.  
RX MEDLINE=90269210; PubMed=2112085;  
RA Herz J., Kowal R.C., Goldstein J.L., Brown M.S.;  
RT "Proteolytic processing of the 600 kd low density lipoprotein  
RT receptor-related protein (LRP) occurs in a trans-Golgi  
RT compartment.";  
RL EMBO J. 9:1769-1776(1990).  
RN [3]  
RP FUNCTION.  
RX MEDLINE=91092405; PubMed=1702392;  
RA Kristensen T., Moestrup S.K., Gliemann J., Bendtsen L., Sand O.,  
RA Sottrup-Jensen L.;  
RT "Evidence that the newly cloned low-density-lipoprotein receptor  
RT related protein (LRP) is the alpha 2-macroglobulin receptor.";  
RL FEBS Lett. 276:151-155(1990).  
RN [4]  
RP STRUCTURE BY NMR OF 1059-1100.  
RX MEDLINE=99253972; PubMed=10318830;  
RA Huang W., Dolmer K., Gettins P.G.W.;  
RT "NMR solution structure of complement-like repeat CR8 from the low  
RT density lipoprotein receptor-related protein.";  
RL J. Biol. Chem. 274:14130-14136(1999).  
CC -!- FUNCTION: INVOLVED IN THE PLASMA CLEARANCE OF CHYLOMICRON REMNANTS  
AND ACTIVATED ALPHA 2-MACROGLOBULIN, AS WELL AS THE LOCAL

CC METABOLISM OF COMPLEXES BETWEEN PLASMINOGEN ACTIVATORS AND THEIR  
CC ENDOGENOUS INHIBITORS.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN LIVER, BRAIN AND LUNG.  
CC -!- PTM: CLEAVED INTO A 85 KDA MEMBRANE-SPANNING SUBUNIT (LRP-85) AND  
CC A 515 KDA LARGE EXTRACELLULAR DOMAIN (LRP-515) THAT REMAINS NON-  
CC COVALENTLY ASSOCIATED.  
CC -!- SIMILARITY: CONTAINS 22 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 31 LDL-RECEPTOR CLASS A DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X13916; CAA32112.1; -.  
DR PDB; 1CR8; 06-JAN-99.  
DR MIM; 107770; -.  
DR InterPro; IPR000033; -.  
DR InterPro; IPR000152; -.  
DR InterPro; IPR000561; -.  
DR InterPro; IPR001881; -.  
DR InterPro; IPR002172; -.  
DR Pfam; PF00008; EGF; 16.  
DR Pfam; PF00057; ldl\_recept\_a; 31.  
DR Pfam; PF00058; ldl\_recept\_b; 33.  
DR PRINTS; PR00261; LDLRECEPTOR.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
DR PROSITE; PS00022; EGF\_1; 5.  
DR PROSITE; PS01186; EGF\_2; 8.  
DR PROSITE; PS01187; EGF\_CA; 2.  
DR PROSITE; PS01209; LDLRA\_1; 27.  
DR PROSITE; PS50068; LDLRA\_2; 31.  
KW Receptor; Transmembrane; Repeat; Endocytosis; Glycoprotein;  
KW Signal; Calcium-binding; EGF-like domain; Coated pits; 3D-structure.  
FT SIGNAL 1 19  
FT CHAIN 20 4544  
FT LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED  
FT PROTEIN 1.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT LDL-RECEPTOR CLASS A 1.  
FT LDL-RECEPTOR CLASS A 2.  
FT EGF-LIKE 1.  
FT EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT EGF-LIKE 3.  
FT EGF-LIKE 4.  
FT LDL-RECEPTOR CLASS A 3.  
FT LDL-RECEPTOR CLASS A 4.  
FT LDL-RECEPTOR CLASS A 5.  
FT LDL-RECEPTOR CLASS A 6.  
FT LDL-RECEPTOR CLASS A 7.  
FT LDL-RECEPTOR CLASS A 8.  
FT LDL-RECEPTOR CLASS A 9.  
FT LDL-RECEPTOR CLASS A 10.  
FT EGF-LIKE 5.  
FT EGF-LIKE 6.  
FT EGF-LIKE 7.  
FT EGF-LIKE 8.  
FT EGF-LIKE 9.  
FT EGF-LIKE 10.  
FT LDL-RECEPTOR CLASS A 11.  
FT LDL-RECEPTOR CLASS A 12.  
FT LDL-RECEPTOR CLASS A 13.  
FT LDL-RECEPTOR CLASS A 14.  
FT LDL-RECEPTOR CLASS A 15.  
FT LDL-RECEPTOR CLASS A 16.  
FT LDL-RECEPTOR CLASS A 17.  
FT LDL-RECEPTOR CLASS A 18.  
FT LDL-RECEPTOR CLASS A 19.

FT	DOMAIN	2902	2940	LDL-RECEPTOR	CLASS A 20.		FT	DISULFID	1227	1237	BY SIMILARITY.
FT	DOMAIN	2941	2981	EGF-LIKE 11.			FT	DISULFID	1233	1246	BY SIMILARITY.
FT	DOMAIN	2982	3022	EGF-LIKE 12.	CALCIUM-BINDING (POTENTIAL).		FT	DISULFID	1248	1261	BY SIMILARITY.
FT	DOMAIN	3290	3331	EGF-LIKE 13.			FT	DISULFID	1540	1553	BY SIMILARITY.
FT	DOMAIN	3332	3371	LDL-RECEPTOR	CLASS A 21.		FT	DISULFID	1549	1563	BY SIMILARITY.
FT	DOMAIN	3372	3410	LDL-RECEPTOR	CLASS A 22.		FT	DISULFID	1565	1578	BY SIMILARITY.
FT	DOMAIN	3411	3450	LDL-RECEPTOR	CLASS A 23.		Query Match 29.8%; Score 43.5; DB 1; Length 4544;				
FT	DOMAIN	3451	3491	LDL-RECEPTOR	CLASS A 24.		Best Local Similarity 41.7%; Pred. No. 7.8e+02; Indels 3; Gaps 1;				
FT	DOMAIN	3492	3533	LDL-RECEPTOR	CLASS A 25.		Matches 10; Conservative 3; Mismatches 8; Indels 3; Gaps 1;				
FT	DOMAIN	3534	3572	LDL-RECEPTOR	CLASS A 26.		QY 6 YQGVPRAPGTEARRH--YDEGV 26				
FT	DOMAIN	3573	3611	LDL-RECEPTOR	CLASS A 27.		Db 3926 YRSLPPAAPPTTSNRHRRQIDRGV 3949				
FT	DOMAIN	3612	3649	LDL-RECEPTOR	CLASS A 28.		RESULT 29				
FT	DOMAIN	3652	3692	LDL-RECEPTOR	CLASS A 29.		D7_XENLA				
FT	DOMAIN	3693	3733	LDL-RECEPTOR	CLASS A 30.		ID D7_XENLA STANDARD; PRT; 278 AA.				
FT	DOMAIN	3739	3778	LDL-RECEPTOR	CLASS A 31.		AC P13007;				
FT	DOMAIN	3781	3823	EGF-LIKE 14.			DT 01-JAN-1990 (Rel. 13; Created)				
FT	DOMAIN	3824	3861	EGF-LIKE 15.			DT 01-JAN-1990 (Rel. 13; Last sequence update)				
FT	DOMAIN	4147	4183	EGF-LIKE 16.			DE D7 PROTEIN.				
FT	DOMAIN	4196	4232	EGF-LIKE 17.			GN Xenopus laevis (African clawed frog).				
FT	DOMAIN	4232	4268	EGF-LIKE 18.			OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
FT	DOMAIN	4268	4304	EGF-LIKE 19.			OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;				
FT	DOMAIN	4304	4340	EGF-LIKE 20.			OC Xenopodinae; Xenopus.				
FT	DOMAIN	4340	4375	EGF-LIKE 21.			OX NCBI_TaxID=8355;				
FT	DOMAIN	4373	4409	EGF-LIKE 22.			RN [1]				
FT	SITE	3940	3943	RECOGNITION SITE FOR PROTEOLYTICAL			RP SEQUENCE FROM N.A.				
FT	SITE	4473	4473	PROCESSING (POTENTIAL).			RC TISSUE-Oocyte;				
FT	DOMAIN	4502	4507	CRITICAL FOR ENDOCYTOSIS (POTENTIAL).			RX MEDLINE=89079004; PubMed=3203907;				
FT	DISULFID	27	40	ENDOCYTOSIS SIGNAL (POTENTIAL).			RA Smith R.C.; Dworkin M.B.; Dworkin-Rastl E.;				
FT	DISULFID	34	53	BY SIMILARITY.			RT "Destruction of a translationally controlled mRNA in Xenopus oocytes				
FT	DISULFID	47	64	BY SIMILARITY.			RL Genes Dev. 2:1296-1306(1988).				
FT	DISULFID	72	85	BY SIMILARITY.			CC -!- FUNCTION: INVOLVED IN OOCYTE MATURATION. IT IS POSSIBLE THAT D7 IS				
FT	DISULFID	79	98	BY SIMILARITY.			CC MATURATION CANNOT PROCEED BEYOND THIS POINT UNLESS A THRESHOLD				
FT	DISULFID	92	108	BY SIMILARITY.			CC AMOUNT OF D7 PROTEIN IS PROVIDED.				
FT	DISULFID	115	124	BY SIMILARITY.			CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.				
FT	DISULFID	120	133	BY SIMILARITY.			CC -!- DEVELOPMENTAL STAGE: ITS LEVELS ARE HIGHEST DURING THE FIRST DAY				
FT	DISULFID	135	148	BY SIMILARITY.			CC OF EMBRYONIC DEVELOPMENT AND THEN DECREASE; D7 PROTEIN WAS NOT				
FT	DISULFID	154	164	BY SIMILARITY.			CC DETECTED IN ADULT TISSUES.				
FT	DISULFID	160	173	BY SIMILARITY.			CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
FT	DISULFID	175	188	BY SIMILARITY.			CC between the Swiss Institute of Bioinformatics and the EMBL outstation -				
FT	DISULFID	175	188	BY SIMILARITY.			CC the European Bioinformatics Institute. There are no restrictions on its				
FT	DISULFID	478	504	BY SIMILARITY.			CC use by non-profit institutions as long as its content is in no way				
FT	DISULFID	489	504	BY SIMILARITY.			CC modified and this statement is not removed. Usage by and for commercial				
FT	DISULFID	506	519	BY SIMILARITY.			CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
FT	DISULFID	807	818	BY SIMILARITY.			CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .				
FT	DISULFID	814	827	BY SIMILARITY.			CC EMBL; X13856; CAA32068.1;				
FT	DISULFID	829	842	BY SIMILARITY.			CC PIR; S06173; S06173.				
FT	DISULFID	854	866	BY SIMILARITY.			CC Developmental protein.				
FT	DISULFID	861	879	BY SIMILARITY.			CC SEQUENCE 278 AA; 31384 MW; 6BD0B882DAAAE4FE CRC64;				
FT	DISULFID	873	890	BY SIMILARITY.			Query Match 29.5%; Score 43; DB 1; Length 278;				
FT	DISULFID	895	907	BY SIMILARITY.			Best Local Similarity 47.1%; Pred. No. 56;				
FT	DISULFID	902	920	BY SIMILARITY.			Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;				
FT	DISULFID	914	931	BY SIMILARITY.			QY 7 QGVPRAPGTEARRHYD 23				
FT	DISULFID	936	948	BY SIMILARITY.			Db 152 QNQEPEPFTSSERNYD 168				
FT	DISULFID	943	961	BY SIMILARITY.			RESULT 30				
FT	DISULFID	955	971	BY SIMILARITY.			CC08_CAEEL				
FT	DISULFID	976	989	BY SIMILARITY.			ID CC08_CAEEL STANDARD; PRT; 282 AA.				
FT	DISULFID	984	1002	BY SIMILARITY.							
FT	DISULFID	996	1011	BY SIMILARITY.							
FT	DISULFID	1015	1027	BY SIMILARITY.							
FT	DISULFID	1022	1040	BY SIMILARITY.							
FT	DISULFID	1034	1051	BY SIMILARITY.							
FT	DISULFID	1062	1075	BY SIMILARITY.							
FT	DISULFID	1069	1088	BY SIMILARITY.							
FT	DISULFID	1082	1097	BY SIMILARITY.							
FT	DISULFID	1104	1118	BY SIMILARITY.							
FT	DISULFID	1112	1131	BY SIMILARITY.							
FT	DISULFID	1125	1140	BY SIMILARITY.							
FT	DISULFID	1145	1159	BY SIMILARITY.							
FT	DISULFID	1152	1172	BY SIMILARITY.							
FT	DISULFID	1166	1182	BY SIMILARITY.							
FT	DISULFID	1185	1196	BY SIMILARITY.							
FT	DISULFID	1192	1206	BY SIMILARITY.							
FT	DISULFID	1208	1221	BY SIMILARITY.							



AC P18833; Q19359;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CUTICLE COLLAGEN 8.  
GN COL-8 OR F11H8.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=89326131; PubMed=2753356;  
RA Cox G.N., Fields C., Kramer J.M., Rosenzweig B., Hirsh D.;  
RT "Sequence comparisons of developmentally regulated collagen genes of  
RT Caenorhabditis elegans.";  
RL Gene 76:331-344(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Menezes S.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
CC LINKS.  
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
CC COLLAGENS. COL-7, COL-8 AND COL-19 BELONGS TO THE SAME GROUP OF  
CC COLLAGEN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M25479; AAA27993.1; -.  
DR EMBL; U40187; AAA81159.1; -.  
DR PIR; JS0168; JS0168.  
DR WormPep; F11H8.3; CE04376.  
DR InterPro; IPR000087; -.  
DR InterPro; IPR002486; -.  
DR Pfam; PF01484; Col\_cuticle\_N; 1.  
DR Pfam; PF01391; Collagen; 2.  
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
FT DOMAIN 95 124 TRIPLE-HELICAL REGION.  
FT DOMAIN 141 269 TRIPLE-HELICAL REGION.  
FT CONFLICT 43 43 T -> R (IN REF. 2).  
FT CONFLICT 142 142 C -> R (IN REF. 2).  
FT CONFLICT 190 190 C -> R (IN REF. 2).  
SQ SEQUENCE 282 AA; 27989 MW; 49ABBF66EF9D3D32 CRC64;  
  
Query Match 29.5%; Score 43; DB 1; Length 282;  
Best Local Similarity 80.0%; Pred. No. 57;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 10 PRAEPGTEAR 19  
|| |||||  
Db 196 PRGPGTEYR 205  
  
RESULT 31  
DAF\_HUMAN  
ID DAF\_HUMAN STANDARD; PRT; 381 AA.  
AC P08174; P09679;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE COMPLEMENT DECAY-ACCELERATING FACTOR PRECURSOR (CD55).  
GN DAF OR CD55.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87115845; PubMed=2433596;  
RA Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,  
RA Nussenzweig V.;  
RT "Cloning of decay-accelerating factor suggests novel use of splicing  
RT to generate two proteins.";  
RL Nature 325:545-549(1987).  
RN [2]  
RP SEQUENCE OF 6-381 FROM N.A. (DAF-2 FORM).  
RX MEDLINE=87175602; PubMed=2436222;  
RA Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,  
RA Leykam J.F., Atkinson J.P., Tykocinski M.L.;  
RT "Cloning and characterization of cDNAs encoding the complete sequence  
RT of decay-accelerating factor of human complement.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).  
RN [3]  
RP SEQUENCE OF 1-100 FROM N.A.  
RX MEDLINE=91271256; PubMed=1711208;  
RA Ewulonu U.K., Ravi L., Medof M.E.;  
RT "Characterization of the decay-accelerating factor gene promoter  
RT region.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).  
RN [4]  
RP SEQUENCE OF 35-46.  
RC TISSUE=Urine;  
RX MEDLINE=91291869; PubMed=1712233;  
RA Nakano Y., Sugita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;  
RT "Isolation of two forms of decay-accelerating factor (DAF) from human  
RT urine.";  
RL Biochim. Biophys. Acta 1074:326-330(1991).  
RN [5]  
RP GPI-ANCHOR.  
RX MEDLINE=91093238; PubMed=1824699;  
RA Moran P., Raab H., Kohr W.J., Caras I.W.;  
RT "Glycophospholipid membrane anchor attachment. Molecular analysis of  
RT the cleavage/attachment site.";  
RL J. Biol. Chem. 266:1250-1257(1991).  
RN [6]  
RP DISULFIDE BONDS IN SUSHI DOMAINS.  
RX MEDLINE=92305034; PubMed=1377029;  
RA Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;  
RT "Complete determination of disulfide bonds localized within the short  
RT consensus repeat units of decay accelerating factor (CD55 antigen).";  
RL Biochim. Biophys. Acta 1116:235-240(1992).  
RN [7]  
RP FUNCTION AS A ECHOVIRUS RECEPTOR.  
RX MEDLINE=95045399; PubMed=7525274;  
RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,  
RA Almond J.W.;  
RT "Decay-accelerating factor CD55 is identified as the receptor for  
RT echovirus 7 using CELICS, a rapid immuno-focal cloning method.";  
RL EMBO J. 13:5070-5074(1994).  
RN [8]  
RP VARIANT BLOOD GROUP DR(A-).  
RX MEDLINE=94325573; PubMed=7519480;  
RA Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,  
RA Ferdman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;  
RT "Molecular basis of reduced or absent expression of  
RT decay-accelerating factor in Cromer blood group phenotypes.";  
RL Blood 84:1276-1282(1994).  
CC -!- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT  
CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT  
CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.  
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES  
CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND



CC -!- FUNCTION: HFLC AND HFLK GOVERN THE STABILITY OF PHAGE LAMBDA CII  
CC PROTEIN AND HAVE BEEN PROPOSED TO ENCODE OR REGULATE A CII-  
CC SPECIFIC PROTEASE.  
CC -!- SUBUNIT: HFLC AND HFLK MAY INTERACT TO FORM A MULTIMERIC COMPLEX.  
CC -!- SIMILARITY: BELONGS TO THE BAND 7 / MEC-2 FAMILY. HFLC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U00005; AAC43400.1; -.  
DR EMBL; U14003; AAA97071.1; -.  
DR EMBL; AB000489; AAC77132.1; -.  
DR PIR; C43653; C43653.  
DR EcoGene; EG10435; hflc.  
DR InterPro; IPR001107; -.  
DR Pfam; PF01145; Band\_7; 1.  
KW Hydrolase; Serine protease; Transmembrane.  
FT TRANSMEM 3 23 POTENTIAL.  
SQ SEQUENCE 334 AA; 37649 MW; FC2B0AAC65D27946 CRC64;

Query Match 29.1%; Score 42.5; DB 1; Length 334;  
Best Local Similarity 44.0%; Pred. No. 79;  
Matches 11; Conservative 3; Mismatches 10; Indels 1; Gaps 1;

Qy 1 VEGGLYQGVPRAPGTEARRHYDEG 25  
| | : | : | | | | | : |  
Db 219 VSEAIYNRM-RAEREAVARRHSQG 242

RESULT 33  
MOXR\_METEX STANDARD; PRT; 341 AA.  
AC P30621;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE MOXR PROTEIN (MXAR PROTEIN).  
GN MOXR OR MXAR.  
OS Methylobacterium extorquens.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Methylobacterium group; Methylobacterium.  
OX NCBI\_TaxID=408;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AM1 / NCIB 9133;  
RX MEDLINE=97151721; PubMed=8997703;  
RA Amaratunga K., Goodwin P.M., O'Connor C.D., Anthony C.;  
RT "The methanol oxidation genes mxaFJGIR (S) ACKLD in Methylobacterium  
RT extorquens.";  
RL FEMS Microbiol. Lett. 146:31-38(1997).  
RN [2]  
RP ERRATUM.  
RX MEDLINE=97306655; PubMed=9163922;  
RA Amaratunga K., Goodwin P.M., O'Connor C.D., Anthony C.;  
RL FEMS Microbiol. Lett. 150:175-177(1997).  
RN [3]  
RP SEQUENCE OF 1-81 FROM N.A.  
RC STRAIN-AM1 / NCIB 9133;  
RX MEDLINE=89350892; PubMed=2504152;  
RA Nunn D.N., Day D., Anthony C.;  
RT "The second subunit of methanol dehydrogenase of Methylobacterium  
RT extorquens AM1.";  
RL Biochem. J. 260:857-862(1989).  
CC -!- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF FORMATION OF ACTIVE  
CC METHANOL DEHYDROGENASE.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: WITH THE CORRESPONDING PROTEIN IN P.DENITRIFICANS.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y07864; CAA69191.1; -.  
DR EMBL; X15792; -; NOT\_ANNOTATED\_CDS.  
KW Methanol utilization; ATP-binding.  
FT NP\_BIND 53 60 ATP (POTENTIAL).  
SQ SEQUENCE 341 AA; 38589 MW; E1C45D1370398D7A CRC64;

Query Match 29.1%; Score 42.5; DB 1; Length 341;  
Best Local Similarity 46.2%; Pred. No. 81;  
Matches 12; Conservative 1; Mismatches 10; Indels 3; Gaps 1;

Qy 1 VEGGLYQGVPRAPGTEARRHYDEG 26  
: | | | | | | | | | |  
Db 94 LGE---DGRPRVPEPGVLRRAEDLSV 116

RESULT 34  
YFQB\_SCHPO STANDARD; PRT; 357 AA.  
ID YFQB\_SCHPO  
AC Q10170; Q9Y7I7;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 40.7 KDA PROTEIN C8E11.11 IN CHROMOSOME I.  
GN SPAC8E11.11 OR SPAC26A3.17C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE OF 1-141 FROM N.A.  
RC STRAIN=972;  
RA McLean J., Harris D., Wood V., Barrell B.G., Rajandream M.A.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE OF 77-357 FROM N.A.  
RC STRAIN=972;  
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: SOME, TO RAT GUANIDINACETATE N-METHYLTRANSFERASE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL021817; CAB40198.1; -.  
DR EMBL; Z69240; CAA93240.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 357 AA; 40709 MW; 5529B8D3B88D91A9 CRC64;

Query Match 29.1%; Score 42.5; DB 1; Length 357;  
Best Local Similarity 37.1%; Pred. No. 84;  
Matches 13; Conservative 3; Mismatches 8; Indels 11; Gaps 2;

Qy 4 GLYQGVPRAE--PGTEARRH-----YDEGVR 27  
| : | : | | | | | | | |  
Db 82 GWNIGIDRANETPGCIARRNLLRLYETIVDAGVR 116







```
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).
GN FLAA OR BB0668.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=212; PubMed=8990312;
RC MEDLINE=97144545; PubMed=8990312;
RA Ge Y., Charon N.;
RT "An unexpected flaA homolog is present and expressed in Borrelia burgdorferi."
RL J. Bacteriol. 179:552-556(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35210 / B31;
RC MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RX Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
RL Nature 390:580-586(1997).
RN [3]
RP SEQUENCE OF 276-341 FROM N.A.
RX STRAIN=212; PubMed=9765799;
RC MEDLINE=98438936; PubMed=9765799;
RA Old I.G., Trueba G.A., Saint-Girons I., Johnson R.C.;
RT "A cheA cheW operon in Borrelia burgdorferi, the agent of Lyme disease."
RL Res. Microbiol. 148:191-200(1997).
RN [4]
RP SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.
CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U62900; AAC44770.1;
CC EMBL: AE001168; AAC67025.1; ALT_INIT.
CC EMBL: X91907; CAA63001.1;
CC TIGR: BB0668;
CC Flagella; Periplasmic; Signal.
KW SIGNAL
FT CHAIN 1 19 POTENTIAL.
FT CONFLICT 20 341 FLAGELLAR FILAMENT OUTER LAYER PROTEIN.
FT CONFLICT 254 255 EQ -> VK (IN REF. 1).
FT CONFLICT 317 318 MP -> IA (IN REF. 1 AND 3).
FT CONFLICT 341 341 AA; 38441 MW; 60B1475E8ADF3451 CRC64;
SQ SEQUENCE
```

```
Query Match 28.8%; Score 42; DB 1; Length 341;
Best Local Similarity 61.5%; Pred. No. 95;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 EGLYQGVPRAPPG 15
DB 23 DGLAEGSKRAEPG 35

RESULT 41
A2AB_BOVIN
ID A2AB_BOVIN
AC 077700;

STANDARD; PRT; 392 AA.
```

```
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE ALPHA-2B ADRENERGIC RECEPTOR (ALPHA-2B ADRENOCEPTOR) (FRAGMENT).
GN ADRA2B.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX Stanhope M.J., Madsen O., Waddell V.G., Cleven G.C., de Jong W.W.;
RA Springer M.S.;
RT "Highly congruent molecular support for a diverse superordinal clade of edemic African mammals."
RC Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: ALPHA-2 ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-INDUCED INHIBITION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y15944; CAA75897.1;
CC InterPro: IPR000207;
CC InterPro: IPR000276;
CC InterPro: IPR002233;
CC Pfam: PF00001; 7tm_1; 1.
CC PRINTS; PR00237; GPCRHHODOPSN.
CC PRINTS; PR00559; ADRENERGCA2BR.
CC PRINTS; PR01103; ADRENERGICR.
CC PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
CC PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
CC G-protein coupled receptor; Transmembrane; Multigene family.
KW NON_TER
FT TRANSMEM 1 25 1 (POTENTIAL).
FT DOMAIN 26 36 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 37 62 2 (POTENTIAL).
FT DOMAIN 63 72 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 73 95 3 (POTENTIAL).
FT DOMAIN 96 117 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 118 140 4 (POTENTIAL).
FT DOMAIN 141 156 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 157 180 5 (POTENTIAL).
FT DOMAIN 181 356 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 357 380 6 (POTENTIAL).
FT DOMAIN 381 >392 EXTRACELLULAR (POTENTIAL).
FT DISULFID 72 151 BY SIMILARITY.
FT DOMAIN 283 295 ASP/GLO-RICH (ACIDIC).
FT NON_TER 392 392
SQ SEQUENCE 392 AA; 42723 MW; 6E014510FF47624E CRC64;
```

```
Query Match 28.8%; Score 42; DB 1; Length 392;
Best Local Similarity 37.5%; Pred. No. 11e+02;
Matches 9; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

OY 2 GEGLYQGVPRAPPGTEARRHYDEG 25
DB 270 GQGOKEGVCGSSPEEEAEEEEG 293

RESULT 42
FTSY_MYCLE
ID FTSY_MYCLE
PRT; 430 AA.
```







---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:36 ; Search time 72.61 Seconds  
(without alignments)  
49.198 Million cell updates/sec

Title: US-09-439-313-566  
Perfect score: 146  
Sequence: 1 VEGGLYQGVPRAPGTEARRHYDEGVR 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

- Database : SPTREMBL\_16.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_unclassified.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	69	47.3	599	5 Q9VSV1	Q9vsv1 drosophila
2	63	43.2	742	4 Q94841	Q94841 homo sapien
3	63	43.2	742	6 Q29397	Q29397 bos taurus
4	63	43.2	742	11 Q9JIS5	Q9jis5 mus musculu
5	55	37.7	533	4 Q9Y2W3	Q9y2w3 homo sapien
6	51	34.9	184	1 Q9HKH4	Q9hkh4 thermoplas
7	51	34.9	241	5 Q9W5I2	Q9w5i2 drosophila
8	50.5	34.6	610	2 Q9I231	Q9i231 pseudomonas
9	49.5	33.9	495	2 O52209	O52209 serratia ma
10	49.5	33.9	665	2 Q48373	Q48373 janthinobac
11	49	33.6	121	5 Q9NHM5	Q9nhm5 plasmodium
12	49	33.6	420	2 Q9RKQ2	Q9rkq2 streptomyc
13	49	33.6	2314	2 O69822	O69822 streptomyc
14	48	32.9	153	2 O86586	O86586 streptomyc
15	48	32.9	356	2 Q9K3R1	Q9k3r1 streptomyc
16	48	32.9	482	10 Q9XFZ5	Q9xfz5 vigna mungo
17	48	32.9	616	2 Q49182	Q49182 mycobacteri
18	48	32.9	729	13 Q9PVF7	Q9pvf7 brachydanio
19	48	32.9	808	2 O70021	O70021 saccharopol

20	47.5	32.5	175	10	Q40848	Q40848 picea glauc
21	47.5	32.5	234	2	Q56154	Q56154 streptomyc
22	47	32.2	127	2	Q9PBJ5	Q9pbj5 xylella fas
23	47	32.2	134	2	Q56434	Q56434 thermus aqu
24	47	32.2	216	2	Q9K3M0	Q9k3m0 streptomyc
25	47	32.2	270	8	Q9MF80	Q9mf80 beta vulgar
26	47	32.2	346	1	Q9V077	Q9v077 pyrococcus
27	47	32.2	484	5	Q9TZQ2	Q9tzq2 drosophila
28	47	32.2	486	5	O96660	O96660 drosophila
29	47	32.2	921	10	O82265	O82265 arabidopsis
30	47	32.2	962	10	Q9LGP20	Q9lgp20 arabidopsis
31	47	32.2	1098	10	Q9XGM9	Q9xgm9 arabidopsis
32	46.5	31.8	110	2	Q9RU55	Q9ru55 deinococcus
33	46.5	31.8	4158	5	Q9N3R9	Q9n3r9 caenorhabdi
34	46	31.5	341	4	Q9UE24	Q9ue24 homo sapien
35	46	31.5	388	14	Q65360	Q65360 orgyia pseu
36	46	31.5	558	4	Q9UDQ4	Q9udq4 homo sapien
37	46	31.5	583	2	Q9R7S4	Q9r7s4 escherichia
38	46	31.5	592	1	Q9UZA5	Q9uza5 pyrococcus
39	46	31.5	618	4	Q9ULU3	Q9ulu3 homo sapien
40	46	31.5	774	2	Q9F584	Q9f584 agrobacteri
41	46	31.5	1741	5	Q19206	Q19206 caenorhabdi
42	45.5	31.2	105	2	Q45395	Q45395 bordetella
43	45.5	31.2	180	5	Q9W077	Q9w077 drosophila
44	45.5	31.2	309	2	Q9L6T0	Q9l6t0 salmonella
45	45.5	31.2	680	2	O66027	O66027 mycobacteri
46	45.5	31.2	714	10	O22273	O22273 arabidopsis
47	45.5	31.2	806	10	O65227	O65227 arabidopsis
48	45.5	31.2	819	10	Q9SHU0	Q9shu0 arabidopsis
49	45.5	31.2	1137	5	Q9V5A6	Q9v5a6 drosophila
50	45	30.8	150	11	Q9R230	Q9r230 rattus norv
51	45	30.8	208	2	Q46486	Q46486 corynebacte
52	45	30.8	210	14	P90273	P90273 human immun
53	45	30.8	390	6	Q9GL19	Q9gl19 cynocephalu
54	45	30.8	424	2	Q9KYY6	Q9kyy6 streptomyc
55	45	30.8	431	2	Q9RLD8	Q9rl8 pseudomonas
56	45	30.8	492	11	O88746	O88746 mus musculu
57	45	30.8	565	2	P76523	P76523 escherichia
58	45	30.8	582	1	Q9HRZ5	Q9hrz5 halobacteri
59	44.5	30.5	142	1	Q9HMN4	Q9hmn4 halobacteri
60	44.5	30.5	177	2	Q9KYR2	Q9kyr2 streptomyc
61	44.5	30.5	186	11	O35906	O35906 mus musculu
62	44.5	30.5	235	1	Q9YFJ4	Q9yjf4 aeropyrum p
63	44.5	30.5	340	10	Q9LGD2	Q9lqd2 oryza sativ
64	44.5	30.5	413	2	P95236	P95236 mycobacteri
65	44	30.1	196	1	Q9Y9E8	Q9y9e8 aeropyrum p

ALIGNMENTS

RESULT	1
Q9VSV1	
ID	Q9VSV1
AC	Q9VSV1;
DT	01-MAY-2000 (TRENBLrel. 13, Created)
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT	01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE	CG4484 PROTEIN.
GN	CG4484.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Chertis J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster";  
RT Science 287:2185-2195(2000).  
RL EMBL; AE003552; AAF50310.1;  
DR FlyBase; FBgn0035968; CG4484.  
SQ SEQUENCE 599 AA; 66057 MW; C5381D334CFBF2EB CRC64;

Query Match 47.3%; Score 69; DB 5; Length 599;

Best Local Similarity 48.1%; Pred. No. 0.13; Length 599;

Matches 13; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 1 VGEGLYQGVPRAPGTEARRHYDEGVR 27  
||| :| | | | :| | :| |  
Db 414 VGEAVFHGDPATAAPNSEAALNYEAGVR 440

RESULT 2

ID O94841 PRELIMINARY; PRT; 742 AA.

AC O94841;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE KIAA0736 PROTEIN.

GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99087487; PubMed=9872452;  
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,  
RA Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XI.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:277-286(1998).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.

DR EMBL; AB018279; BAA34456.1;  
DR InterPro; IPR001066;  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 742 AA; 82695 MW; 913E216D5CFC2FB2 CRC64;

Query Match 43.2%; Score 63; DB 4; Length 742;  
Best Local Similarity 51.7%; Pred. No. 1.1;  
Matches 15; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

OY 3 EGLYQGVPRAPGTEARRHYD----EGVR 27  
||| :| | | | :| | :| |  
Db 95 EGEYQGIPTAESGGKGERMADGAPLAGVR 123

RESULT 3

ID Q29397 PRELIMINARY; PRT; 742 AA.

AC Q29397;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE P87-TRANSPORTER-LIKE PROTEIN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93050176; PubMed=1426240;  
RA Gingsrich J.A., Andersen P.H., Tiberi M., el Mestikawy S.,  
RA Jorgensen P.N., Freneau R.T. Jr., Caron M.G.;  
RT "Identification, characterization, and molecular cloning of a novel  
RT transporter-like protein localized to the central nervous system.";  
RL FEBS Lett. 312:115-122(1992).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL; S47919; AAB24028.1;  
DR InterPro; IPR001066;  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 742 AA; 82589 MW; 58AE0C9BB224DEB8 CRC64;

Query Match 43.2%; Score 63; DB 6; Length 742;

Best Local Similarity 51.7%; Pred. No. 1.1;

Matches 15; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

OY 3 EGLYQGVPRAPGTEARRHYD----EGVR 27  
||| :| | | | :| | :| |  
Db 95 EGEYQGIPTAESGGKGERMADGAPLAGVR 123

RESULT 4

ID Q9JIS5 PRELIMINARY; PRT; 742 AA.

AC Q9JIS5;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CA2+ REGULATOR SV2A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RX MEDLINE=20088300; PubMed=10624962;

RA Janz R., Goda Y., Geppert M., Missler M., Sudhof T.C.;  
RT "SV2A and SV2B function as redundant Ca2+ regulators in  
RT neurotransmitter release";  
RL Neuron 24:1003-1016(1999).  
DR EMBL; AF196781; AAF87321.1; -.  
DR EMBL; AF196780; AAF87321.1; JOINED.  
DR InterPro; IPR001066; -.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN.1.  
SQ SEQUENCE 742 AA; 82647 MW; 1074857FD13ED894 CRC64;

Query Match 43.2%; Score 63; DB 11; Length 742;  
Best Local Similarity 51.7%; Pred. No. 1.1;  
Matches 15; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 3 EGLYQGVPRAPGTEARRHYD----EGVR 27  
|||:|||||:|:|  
Db 95 EGEYQGIPIRAESGGKGERMADGAPLAGVR 123

RESULT 5  
QY2W3  
ID Q9Y2W3 PRELIMINARY; PRT; 533 AA.  
AC Q9Y2W3;

DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE DNB-5 (FRAGMENT).

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Amler L.C., Bauer A., Corvi R., Dihlmann S., Praml C., Cavenee W.K.,  
RA Schwab M., Hampton G.M.;  
RT "Identification and Characterization of Novel Genes Located at the  
RT t(1;15)(p36.2;q24) Translocation Breakpoint in the Neuroblastoma Cell  
RT Line NGP";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF118274; AAD27583.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 533 AA; 57048 MW; 6AF8BA162FC9D1C8 CRC64;

Query Match 37.7%; Score 55; DB 4; Length 533;  
Best Local Similarity 42.3%; Pred. No. 11;  
Matches 11; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VGEGLYQGVPRAPGTEARRHYDEGV 26  
:||:|||||:|:|:|:|  
Db 333 MGEVVFQDGPAPHTSEAYQKNSGV 358

RESULT 6  
Q9HKH4  
ID Q9HKH4 PRELIMINARY; PRT; 184 AA.  
AC Q9HKH4;

DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL PROTEIN TA0625.  
GN TA0625.

OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;  
OC Thermoplasma.  
OX NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1728;  
RX MEDLINE=20479972; PubMed=11029001;  
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,

RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
RT acidophilum";  
RL Nature 407:508-513(2000).  
DR EMBL; AL445064; CAC11764.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 184 AA; 20720 MW; 6C416E6059543D0F CRC64;

Query Match 34.9%; Score 51; DB 1; Length 184;  
Best Local Similarity 36.0%; Pred. No. 13;  
Matches 9; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 3 EGLYQGVPRAPGTEARRHYDEGVR 27  
||:||||:|:|:|:|  
Db 11 EGRWQGINSDIPEESLRHFEER 35

RESULT 7  
Q9W5I2  
ID Q9W5I2 PRELIMINARY; PRT; 241 AA.  
AC Q9W5I2;

DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE CG17683 PROTEIN.  
GN CG17683.

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,  
RA Borkova D., Botchan M.R., Bouck J., Butler H., Cadieu E., Center A., Chandra I.,  
RA Burtis K.C., Busam D.A., Butler H., Dahlke C., Davenport L.B., Davies P.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Jalali M., Kalush F., Karpen G.H., Kravitz S., Kulp D., Lai Z.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Liang Y., Lin X.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
\*The genome sequence of Drosophila melanogaster.\*;



RL Science 287:2185-2195(2000).  
DR EMBL: AF003231; AAF45404.1; --  
DR FlyBase: FBgn0040002; CGI7683.  
SQ SEQUENCE 241 AA; 28118 MW; A5E035EEA91BEDBF CRC64;

Query Match 34.9%; Score 51; DB 5; Length 241;  
Best Local Similarity 47.4%; Pred. No. 17;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 EGYQGVPRAPGTEARRH 21  
| | | | | : | | | | |  
Db 207 EELYQNLPSEPENSLTKH 225

RESULT 8  
Q9I231 ID Q9I231 PRELIMINARY; PRT; 610 AA.  
AC Q9I231;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PROBABLE ASPARAGINE SYNTHETASE.  
GN PA2084.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.; an  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL: AE004636; AAG05472.1; --  
DR InterPro; IPR00583; --  
DR InterPro; IPR001962; --  
DR Pfam; PF00310; GARase\_2; 1.  
DR Pfam; PF00733; Asn\_synthase; 2.  
SQ SEQUENCE 610 AA; 68024 MW; 9150900739043AC6 CRC64;

Query Match 34.6%; Score 50.5; DB 2; Length 610;  
Best Local Similarity 46.2%; Pred. No. 53;  
Matches 12; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 2 GEGLYQGVPRAPGTEARRHYDEGVR 27  
| | | | | : | | | | |  
Db 194 GHGLYKGLRQVRPG-QALRFRNGVR 218

RESULT 9  
O52209 ID O52209 PRELIMINARY; PRT; 495 AA.  
AC O52209;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE RETA.  
GN RETA.  
OS Serratia marcescens.  
OC Plasmid R471a.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Serratia.  
OX NCBI\_TaxID=615;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=75096689; PubMed=1089756;  
RA Hedges R.W., Rodriguez-Lemoine V., Datta N.;  
RT "R factors from Serratia marcescens.";  
RL J. Gen. Microbiol. 86:88-92(1975).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93374835; PubMed=8366028;  
RA HO C., Kulaeva O.I., Levine A.S., Woodgate R.;  
RT "A rapid method for cloning mutagenic DNA repair genes: isolation of  
RT umu-complementing genes from multidrug resistance plasmids R391,  
RT R446b, and R471a.";  
RL J. Bacteriol. 175:5411-5419(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kulaeva O.I., Koonin E.V., Wootton J.C., Levine A.S., Woodgate R.;  
RL Mutat. Res. 0:0-0(1997).  
CC -!- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE  
CC TRANSCRIPTASE).  
DR EMBL: AF027768; AAC82519.1; --  
DR InterPro; IPR000477; --  
DR Pfam; PF00078; xvt; 1.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 495 AA; 57746 MW; ECA93A980D961A15 CRC64;

Query Match 33.9%; Score 49.5; DB 2; Length 495;  
Best Local Similarity 61.1%; Pred. No. 59;  
Matches 11; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 5 LYQGVPRAPGTEARRHY 22  
| | | | | : | | | | |  
Db 120 LHQGSYRAQPG---RRHY 134

RESULT 10  
Q48373 ID Q48373 PRELIMINARY; PRT; 665 AA.  
AC Q48373;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CHITINASE PRECURSOR.  
GN CHI 69.  
OS Janthinobacterium lividum.  
OC Bacteria; Proteobacteria; beta subdivision; Oxalobacter group;  
OC Janthinobacterium.  
OX NCBI\_TaxID=29581;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96013069; PubMed=7557339;  
RA Gleave A.P., Taylor R.K., Morris B.A., Greenwood D.R.;  
RT "Cloning and sequencing of a gene encoding the 69-kDa extracellular  
RT chitinase of Janthinobacterium lividum.";  
RL FEMS Microbiol. Lett. 131:279-288(1995).  
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
CC HYDROLASES).  
DR EMBL: U07025; AAA83223.1; --  
DR HSSP; P07254; ICTN.  
DR InterPro; IPR00508; --  
DR InterPro; IPR001223; --  
DR InterPro; IPR001579; --  
DR InterPro; IPR003610; --  
DR Pfam; PF00704; Glyco\_hydro\_18; 2.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
DR PROSITE; PS00761; SPASE\_1\_3; UNKNOWN\_1.  
DR SMART; SM00495; ChtBD3; 1.  
KW Glycosidase; Hydrolase; Signal.  
FT SIGNAL 23 POTENTIAL.  
FT CHAIN 24 665  
SQ SEQUENCE 665 AA; 69748 MW; FCF371332C1C6839 CRC64;

Query Match 33.9%; Score 49.5; DB 2; Length 665;  
Best Local Similarity 40.0%; Pred. No. 81;  
Matches 10; Conservative 3; Mismatches 7; Indels 5; Gaps 1;

QY 2 GEGLYQGVPRAPGTEARRHYDEGV 26  
|:||||| ||| |:|:  
Db 569 GDGLYQAATGAPGT-----YESGI 588

RESULT 11  
Q9NHM5 PRELIMINARY; PRT; 121 AA.  
AC Q9NHM5;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE PFEMP1 PROTEIN (FRAGMENT).  
GN VAR.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Taylor H.M., Newbold C.I.;  
RT "Var gene diversity in Plasmodium falciparum is generated by frequent recombination events.";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF221765; AAF36605.1; -.  
FT NON\_TER 1  
FT NON\_TER 121  
SQ SEQUENCE 121 AA; 14390 MW; 3A2E83436826AA2A CRC64;

Query Match 33.6%; Score 49; DB 5; Length 121;  
Best Local Similarity 33.3%; Pred. No. 15;  
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 LYQGVPRAPGTEARRHYDEG 25  
|:|:|:|:|:|:|:|:|:|:  
Db 33 IYEELKRGDKTNEAEKHEDG 53

RESULT 12  
Q9RKQ2 PRELIMINARY; PRT; 420 AA.  
AC Q9RKQ2;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE PUTATIVE MEMBRANE PROTEIN.  
GN SCC75A.22.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]

Query Match 33.6%; Score 49; DB 2; Length 2314;  
Best Local Similarity 43.5%; Pred. No. 3.6e+02;  
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;  
QY 2 GEGLYQGVPRAPGTEARRHYDE 24  
|:|:|:|:|:|:|:|:|:|:  
Db 486 GTQLRQGYIAGDPGTTVQRGYDQ 508  
RESULT 14  
O86586

RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL133220; CAB61722.1; -.  
SQ SEQUENCE 420 AA; 44522 MW; 563444615594E01BC CRC64;

Query Match 33.6%; Score 49; DB 2; Length 420;  
Best Local Similarity 64.3%; Pred. No. 58;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VEGELYQGVPRAP 14  
|:|:|:|:|:|:|:|:|:|:  
Db 111 VEGAYGGLPEAPP 124  
PRELIMINARY; PRT; 2314 AA.

RESULT 13  
O69822 PRELIMINARY; PRT; 2314 AA.  
AC O69822;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PUTATIVE SECRETED PROTEIN.  
GN SCIA6.17C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]

Query Match 33.6%; Score 49; DB 2; Length 2314;  
Best Local Similarity 43.5%; Pred. No. 3.6e+02;  
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;  
QY 2 GEGLYQGVPRAPGTEARRHYDE 24  
|:|:|:|:|:|:|:|:|:|:  
Db 486 GTQLRQGYIAGDPGTTVQRGYDQ 508  
RESULT 14  
O86586

Query Match 33.6%; Score 49; DB 2; Length 2314;  
Best Local Similarity 43.5%; Pred. No. 3.6e+02;  
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEARRHYDE 24  
|:|:|:|:|:|:|:|:|:|:  
Db 486 GTQLRQGYIAGDPGTTVQRGYDQ 508

RESULT 14  
O86586



RT "Cloning and DNA sequence of the Mycobacterium fortuitum var fortuitum  
RT plasmid pAL5000.";  
RL Plasmid 27:130-140(1992).  
DR EMBL; M60875; AAA25373.1; -.  
DR InterPro; IPR000345; -.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN1.  
SQ SEQUENCE 616 AA; 66784 MW; 715EEBE768D9A856 CRC64;

Query Match 32.9%; Score 48; DB 2; Length 616;  
Best Local Similarity 64.3%; Pred. No. 1.2e+02;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 12 AEPGTEARRHYDEG 25  
Db 151 AHPGTEPDRHYRNG 164

RESULT 18  
Q9PVF7 PRELIMINARY; PRT; 729 AA.  
AC Q9PVF7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE CELL-ADHESION PROTEIN PLAKOGLOBIN.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99386700; PubMed=10456847;  
RA Cerda J., Reidenbach S., Pratzel S., Franke W.W.;  
RT "Cadherin-catenin complexes during zebrafish oogenesis: heterotypic  
RT junctions between oocytes and follicle cells.";  
RL Biol. Reprod. 61:692-704(1999).  
DR EMBL; AF099738; AAD56592.1; -.  
DR HSSP; Q02248; LDOW.  
DR InterPro; IPR000225; -.  
DR Pfam; PF00514; Armadillo\_seg; 11.  
DR PROSITE; PS50176; ARM\_REPEAT; 8.  
DR SMART; SM00185; ARM; 1.  
SQ SEQUENCE 729 AA; 80033 MW; 91E00417B4FD8CEE CRC64;

Query Match 32.9%; Score 48; DB 13; Length 729;  
Best Local Similarity 53.3%; Pred. No. 1.5e+02;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 13 EPGTEARRHYDEGVR 27  
Db 529 KPGSSAQRSYQDGVR 543

RESULT 19  
O70021 PRELIMINARY; PRT; 808 AA.  
AC O70021;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE BETA-GLUCOSIDASE.  
GN ERYBI.  
OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;  
OC Saccharopolyspora.  
OX NCBI\_TaxID=1836;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL2338;

RA Gaiser S., Boehm G.A., Doumith M., Raynal M.C., Dhillon N.,  
RA Cortes J., Leadlay P.F.;  
RL Mol. Gen. Genet. 0:0-0(0).  
DR EMBL; Y14327; CAA74702.1; -.  
DR InterPro; IPR001764; -.  
DR InterPro; IPR002772; -.  
DR Pfam; PF00933; Glyco\_hydro\_3; 1.  
DR Pfam; PF01915; Glyco\_hydro\_3\_C; 1.  
DR PRINTS; PR00133; GLHYDRLASE3.  
SQ SEQUENCE 808 AA; 86677 MW; 0C619016CC00A751 CRC64;

Query Match 32.9%; Score 48; DB 2; Length 808;  
Best Local Similarity 42.1%; Pred. No. 1.6e+02;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 8 GVPRAEPGTEARRHYDEGV 26  
Db 651 GDPARFPGVDGKVHYSEGI 669

RESULT 20  
Q40848 PRELIMINARY; PRT; 175 AA.  
AC Q40848;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)  
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN.  
GN EMB23.  
OS Picea glauca (White spruce).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Coniferopsida; Coniferales; Pinaceae; Picea.  
OX NCBI\_TaxID=3330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dong J.Z., Dunstan D.I.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; L47603; AAB01558.1; -.  
DR Mendel; 12550; Picgl; 1750; 12550.  
SQ SEQUENCE 175 AA; 19128 MW; 0CD8BD4B3F80BA0B CRC64;

Query Match 32.5%; Score 47.5; DB 10; Length 175;  
Best Local Similarity 48.0%; Pred. No. 37;  
Matches 12; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

Qy 1 VEGELYQGVPR-AEPGTEARRHYDE 24  
Db 114 LSEGAEATVERDADPHVEVRREYDE 138

RESULT 21  
Q56154 PRELIMINARY; PRT; 234 AA.  
AC Q56154;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE JADR1.  
GN JADR1.  
OS Streptomyces violaceus (Streptomyces venezuelae).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1936;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISP5230;  
RX MEDLINE=96042086; PubMed=7592375;  
RA Yang K., Han L., Vining L.C.;  
RT "Regulation of jadomycin B production in Streptomyces venezuelae  
RT ISP5230: involvement of a repressor gene, jadR2.";  
RL J. Bacteriol. 177:6111-6117(1995).



CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER  
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.  
CC -!- SIMILARITY: TO BACTERIAL REGULATORY PROTEINS INVOLVED IN SIGNAL  
CC TRANSDUCTION.  
DR EMBL; U24659; AAB36584.1; --  
DR HSSP; P08402; 1B00.  
DR InterPro; IPR001789; --  
DR InterPro; IPR001867; --  
DR Pfam; PF00072; response\_reg; 1.  
DR Pfam; PF00486; trans\_reg\_C; 1.  
DR SMART; SM00448; REC; 1.  
KW DNA-binding; Phosphorylation; Sensory transduction;  
KW Transcription regulation;  
SQ SEQUENCE 234 AA; 25606 MW; CC251108FEC39D6D CRC64;

Query Match 32.5%; Score 47.5; DB 2; Length 234;  
Best Local Similarity 43.3%; Pred. NO. 51;  
Matches 13; Conservative 2; Mismatches 8; Indels 7; Gaps 1;

OY 2 GEGLYQGVPRAEPTGTEARRHYDE 24  
|||||  
DB 30 GEGLYQGVPRAEPTGTEARRHYDE 24  
|||||

RESULT 22  
Q9PBJ5 PRELIMINARY; PRT; 127 AA.  
AC Q9PBJ5;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE APAG PROTEIN.  
GN XF2149.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_TaxID=2371;

[1]  
SEQUENCE FROM N.A.  
STRAIN-9A5C;  
MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Barrios M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RT "The genome sequence of the plant pathogen xylella fastidiosa";  
Nature 406:151-159(2000).  
DR EMBL; AE004029; AAF84948.1; --  
SQ SEQUENCE 127 AA; 14247 MW; 3DD5206C2937BE25 CRC64;

Query Match 32.2%; Score 47; DB 2; Length 127;  
Best Local Similarity 48.0%; Pred. NO. 31;  
Matches 12; Conservative 1; Mismatches 10; Indels 2; Gaps 1;

OY 2 GEGLYQGVPRAEPTGTEARRHYDEGV 26  
|||||  
DB 66 GEGLYQGVPRAEPTGTEARRHYDEGV 26  
|||||

RESULT 23  
Q56434 PRELIMINARY; PRT; 134 AA.  
AC Q56434;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE RIBOSOMAL PROTEIN L11 (FRAGMENT).  
GN RPL11.  
OS Thermus aquaticus (subsp. thermophilus).  
OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.  
OX NCBI\_TaxID=274;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94124036; PubMed=8294036;  
RA Heinrich T., Erdmann V.A., Hartmann R.K.;  
RT "Sequence of the gene encoding ribosomal protein L11 from Thermus  
RL thermophilus HB8.";  
DR EMBL; L10371; AAA27503.1; --  
KW Ribosomal protein.  
FT NON\_TER  
SQ SEQUENCE 134 AA; 14336 MW; C8EB750B7B75E5FFC CRC64;

Query Match 32.2%; Score 47; DB 2; Length 134;  
Best Local Similarity 41.7%; Pred. NO. 33;  
Matches 10; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

OY 2 GEGLYQGVPRAEPTGTEARRHYDEG 25  
|||||  
DB 40 GRGLGEGGQARPGEGPDHLGAG 63  
|||||

RESULT 24  
Q9K3M0 PRELIMINARY; PRT; 216 AA.  
AC Q9K3M0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE PUTATIVE DNA REPAIR PROTEIN.  
GN SCG20A.20C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL360055; CAB96025.1; -.  
SQ SEQUENCE 216 AA; 23424 MW; 15153ACD08419CCA CRC64;

Query Match 32.2%; Score 47; DB 2; Length 216;  
Best Local Similarity 72.7%; Pred. No. 55;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGPT 16  
| | | | | | | |  
Db 186 YHGVPRVHPGT 196

RESULT 25  
Q9MF80  
ID Q9MF80 PRELIMINARY; PRT; 270 AA.  
AC Q9MF80;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE ORF270 PROTEIN.  
GN ORF270.  
OS Beta vulgaris (Sugar beet).  
OG Mitochondrion.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;  
OC Caryophyllales; Chenopodiaceae; Beta.  
OX NCBI\_TaxID=3555;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TK81-O;  
RX MEDLINE=20330382; PubMed=10871408;  
RA Kubo T., Nishizawa S., Sugawara A., Itchoda N., Estiati A., Mikami T.;  
RT "The complete nucleotide sequence of the mitochondrial genome of sugar  
RT beet (Beta vulgaris L.) reveals a novel gene for tRNACys(GCA).";  
RL Nucleic Acids Res. 28:2571-2576(2000).  
DR EMBL; AP00397; BAA99454.1; -.  
KW Mitochondrion.  
SQ SEQUENCE 270 AA; 31030 MW; 0D00E97A90AC902A CRC64;

Query Match 32.2%; Score 47; DB 8; Length 270;  
Best Local Similarity 40.7%; Pred. No. 70;  
Matches 11; Conservative 5; Mismatches 7; Indels 4; Gaps 2;

QY 4 GLYQGV-PRAEPGTEA---RRHYDEGV 26  
| | | : | : | | | | | : | : | :  
Db 122 GLYSNLPQNEPGEYAGIVRAHFDQAI 148

RESULT 26  
Q9V077  
ID Q9V077 PRELIMINARY; PRT; 346 AA.  
AC Q9V077;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE BIFUNCTIONAL SHORT CHAIN ISOPRENYL DIPHOSPHATE SYNTHASE (IDSA).  
GN PAB2389.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORSAY;  
RA Heilig R.;  
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
RT structure and evolution.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ248285; CAB49828.1; -.

DR InterPro; IPR000092; -.  
DR Pfam; PF00348; polyprenyl\_synt; 1.  
DR PROSITE; PS00723; POLYPRENYL\_SYNTHET\_1; 1.  
DR PROSITE; PS00444; POLYPRENYL\_SYNTHET\_2; 1.  
SQ SEQUENCE 346 AA; 38824 MW; 11CEA240E5134F7F CRC64;

Query Match 32.2%; Score 47; DB 1; Length 346;  
Best Local Similarity 45.8%; Pred. No. 91;  
Matches 11; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY 1 VGEGLYQGVPRAPGPT--EARRHY 22  
| | | : | | | | | | | |  
Db 22 VDEKLFELIPEKEPKVLYEAARHY 45

RESULT 27  
Q9TZQ2  
ID Q9TZQ2 PRELIMINARY; PRT; 484 AA.  
AC Q9TZQ2;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE SMOX.  
GN SMOX OR CG2262.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9032828; PubMed=9813169;  
RA Henderson K.D., Andrew D.J.;  
RT "Identification of a novel Drosophila SMAD on the X chromosome.";  
RL Biochem. Biophys. Res. Commun. 252:195-201(1998).  
DR EMBL; AF078529; AAC83344.1; -.  
DR FlyBase; FBgn0025800; SMOX.  
DR InterPro; IPR001132; -.  
DR InterPro; IPR003619; -.  
DR Pfam; PF00968; Dwarf; 1.  
DR SMART; SM00523; DWA; 1.  
SQ SEQUENCE 484 AA; 53475 MW; 9558C5C61A8EC812 CRC64;

Query Match 32.2%; Score 47; DB 5; Length 484;  
Best Local Similarity 45.8%; Pred. No. 1.3e+02;  
Matches 11; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 4 GLYQGVPRAPGTEARRHYDEGVR 27  
| | | | | : | | | : | | |  
Db 331 GLLSNVNRNEVEQTRRHIGKGV 354

RESULT 28  
O96660  
ID O96660 PRELIMINARY; PRT; 486 AA.  
AC O96660;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE SMOX PROTEIN.  
GN SMOX OR SMAD2 OR CG2262.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,



RA Theologis;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AC021198; AAF79886.1; -.  
SQ SEQUENCE 962 AA; 108167 MW; 0454713CBD77280A CRC64;

Query Match 32.2%; Score 47; DB 10; Length 962;  
Best Local Similarity 70.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEGGLYQGVP 10  
:|:|:|:|:|  
Db 820 IGDGLYEGVP 829

RESULT 31  
Q9XGM9 PRELIMINARY; PRT; 1098 AA.  
ID Q9XGM9  
AC Q9XGM9;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE STROMALIN PROTEIN.  
GN SA.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Valdeolmillos A., Barbero J.;  
RT "Molecular Cloning and expression of stromalin protein from  
RT Arabidopsis thaliana.";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AJ242965; CAB45374.1; -.  
SQ SEQUENCE 1098 AA; 125758 MW; 293ACEA3D335471A CRC64;

Query Match 32.2%; Score 47; DB 10; Length 1098;  
Best Local Similarity 50.0%; Pred. No. 3.1e+02;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 QGVPRAEPTGTEARRHYDE 24  
:|:|:|:|:|:|:|:|:|:|  
Db 1034 RGRPRKRPTERKRLFDE 1051

RESULT 32  
Q9RU55 PRELIMINARY; PRT; 110 AA.  
ID Q9RU55  
AC Q9RU55;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 11.3 KDA PROTEIN.  
GN DR1539.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1.";

RL Science 286:1571-1577(1999).  
DR EMBL; AE001998; AAF11112.1; -.  
DR TIGR; DR1539; -.  
KW Hypothetical protein.  
SQ SEQUENCE 110 AA; 11275 MW; D004935674B62BCE CRC64;

Query Match 31.8%; Score 46.5; DB 2; Length 110;  
Best Local Similarity 46.4%; Pred. No. 31;  
Matches 13; Conservative 2; Mismatches 10; Indels 3; Gaps 1;

QY 3 EGLYQGVPRAEPTGTEARRHYDEGVR 27  
|:|:|:|:|:|:|:|:|:|:|  
Db 75 EKVKGAVDRAKAGIHNAEAHASYDEGHR 102

RESULT 33  
Q9N3R9 PRELIMINARY; PRT; 4158 AA.  
ID Q9N3R9  
AC Q9N3R9;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE Y47G6A.23 PROTEIN.  
GN Y47G6A.23.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Minx P., Graves T., Hawrysko C.;  
RT "The sequence of C. elegans cosmid Y47G6A.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AC024791; AAF60650.1; -.  
SQ SEQUENCE 4158 AA; 469192 MW; 3A7A285E107F705D CRC64;

Query Match 31.8%; Score 46.5; DB 5; Length 4158;  
Best Local Similarity 42.3%; Pred. No. 1.5e+03;  
Matches 11; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

QY 3 EGLY---QGVPRAEPTGTEARRHYDEG 25  
:|:|:|:|:|:|:|:|:|:|  
Db 2482 DGLYQAEQGVSSQKPSSTDKKLIYREG 2507

RESULT 34  
Q9UE24 PRELIMINARY; PRT; 341 AA.  
ID Q9UE24  
AC Q9UE24;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE MRNA 3'-REGION (UNKNOWN FUNCTION) (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;





Q9UZA5  
ID Q9UZA5 PRELIMINARY; PRT; 592 AA.  
AC Q9UZA5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE CARBON STARVATION PROTEIN A.  
GN PAB1554.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORSAY;  
RA Heilig R.;  
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";  
RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AJ248287; CAB50154.1; -.  
DR InterPro; IPR001202; -.  
DR PROSITE; PS01159; WW\_DOMAIN\_1; UNKNOWN\_1.  
SQ SEQUENCE 592 AA; 64396 MW; E4B35BB9B6561638 CRC64;  
  
Query Match 31.5%; Score 46; DB 1; Length 592;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 12; Conservative 2; Mismatches 8; Indels 2; Gaps 1;  
  
QY 2 GEGLYQGVPRAPG--TEARRHYD 23  
I::I I::I I::I I::I I::I  
Db 22 GKGLQNKVVRADPNRPTPAHRLYD 45  
  
RESULT 39  
Q9ULU3  
ID Q9ULU3 PRELIMINARY; PRT; 618 AA.  
AC Q9ULU3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE KIAA1126 PROTEIN (FRAGMENT).  
GN KIAA1126.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=20039618; PubMed=10574461;  
RA Hirose M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;  
RT "Characterization of cDNA clones selected by the Genemark analysis from size-fractionated cDNA libraries from human brain.";  
RL DNA Res. 6:329-336(1999).  
DR EMBL; AB032952; BAA86440.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 618 AA; 67578 MW; 312EA4EE31C0E5C4 CRC64;  
  
Query Match 31.5%; Score 46; DB 4; Length 618;  
Best Local Similarity 29.6%; Pred. No. 2.3e+02;  
Matches 8; Conservative 10; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 VGEGLYQGVPRAPGTEARRHYDEGVR 27  
I::I I::I I::I I::I I::I I::I  
Db 383 MGQVIFEGDPKAPSNSTAWQAYNAGVK 409  
  
RESULT 40  
Q9F584  
ID Q9F584 PRELIMINARY; PRT; 774 AA.  
AC Q9F584;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE RIORF171 PROTEIN.  
GN RIORF171.  
OS Agrobacterium rhizogenes.  
OG Plasmid pRi1724.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;  
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N., Yoshida K.;  
RT "The complete nucleotide sequence of a Ri (root inducing) plasmid indicates its chimerical structure between Ti and Sym plasmids.";  
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;  
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;  
RT "Analysis of unique variable region of a plant root inducing plasmid, pRi1724, by the construction of its physical map and library.";  
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RL [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;  
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;  
RT "Genome structure of Ri plasmid (1): Construction of linking library and physical map of pRi1724 in Japanese Agrobacterium.";  
RT Nucleic Acids Symp. Ser. 39:189-190(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF03-01724;  
RX MEDLINE=20241294; PubMed=10780382;  
RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N., Yoshida K.;  
RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and its flanking regions of pRi1724 in Japanese Agrobacterium rhizogenes.";  
RL Nucleic Acids Symp. Ser. 42:67-68(1999).  
DR EMBL; AP002086; BAB16290.1; -.  
KW Plasmid.  
SQ SEQUENCE 774 AA; 86002 MW; 8BBC516E30E8AD01 CRC64;  
  
Query Match 31.5%; Score 46; DB 2; Length 774;  
Best Local Similarity 45.2%; Pred. No. 3e+02;  
Matches 14; Conservative 3; Mismatches 6; Indels 8; Gaps 2;  
  
QY 1 VGEG-----LYQGVPRAPGTEARRHYDEGV 26  
I::I I::I I::I I::I I::I I::I  
Db 43 VGESTSTFVYEGVPLAE---GERRAYEEVW 70  
  
RESULT 41  
Q19206  
ID Q19206 PRELIMINARY; PRT; 1741 AA.  
AC Q19206;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE CODED FOR BY C. ELEGANS CDNA CM21C7.  
GN F08F8.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,



ID Q9L6T0 PRELIMINARY; PRT; 309 AA.  
AC Q9L6T0;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE S. TYPHIMURIUM BRANCHED-CHAIN-AMINO-ACID TRANSAMINASE (ILVE)  
DE (SP:P15168).  
GN ILVE.  
OS Salmonella typhimurium LT2.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=99287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SGSC1412;  
RA WashU;  
RT "The Salmonella typhimurium Genome Sequencing Project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SGSC1412;  
RA Waterston R.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF233324; AAF33481.1; -.  
DR InterPro; IPR001544; -.  
DR Pfam; PF01063; aminotran\_4; 1.  
DR PROSITE; PS00770; AA\_TRANSFER\_CLASS\_4; 1.  
SQ SEQUENCE 309 AA; 34052 MW; 4F60941F3D63CC69 CRC64;

Query Match 31.2%; Score 45.5; DB 2; Length 309;  
Best Local Similarity 31.4%; Pred. No. 1.3e+02;  
Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;

QY 3 EGLYQGV-----RAEP-----GTEARRH-YDEGV 26  
| | | | |  
Db 134 EALDQGIDAMVSSWNRAAPTPTAAKAGGNYLSSLLVGSSEARRHGYQEGI 184  
| | | | |

RESULT 45  
O66027  
ID O66027 PRELIMINARY; PRT; 680 AA.  
AC O66027;  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE METAL TRANSPORTING ATPASE MTA72.  
GN MTA7.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ERDMAN, ATCC 35801;  
RX MEDLINE=98181048; PubMed=9514635;  
RA Calder K.M., Horwitz M.A.;  
RT "Identification of iron-regulated proteins of Mycobacterium  
tuberculosis and cloning of tandem genes encoding a low iron-induced  
protein and a metal transporting ATPase with similarities to two-  
component metal transport systems."  
RL Microb. Pathog. 24:133-143(1998).  
DR EMBL; U82820; AAC15948.1; -.  
DR InterPro; IPR000150; -.  
DR InterPro; IPR001454; -.  
DR InterPro; IPR001757; -.  
DR Pfam; PF00122; E1-E2\_ATPase; 1.  
DR Pfam; PF00702; Hydrolase; 1.  
DR PRINTS; PR00119; CATATPASE.  
DR PROSITE; PS00154; ATPASE\_E1\_E2; UNKNOWN\_1.  
DR PROSITE; PS01229; COF\_2; UNKNOWN\_1.  
SQ SEQUENCE 680 AA; 72149 MW; 0368D58CB53F30BD CRC64;

Query Match 31.2%; Score 45.5; DB 2; Length 680;  
Best Local Similarity 44.0%; Pred. No. 3e+02;  
Matches 11; Conservative 2; Mismatches 11; Indels 1; Gaps 1;  
QY 2 GEGLYQGV-RAEPGTEARRHYDEG 25  
| | | | | | | | | | | | | | | | |  
Db 67 GTGLARGAPLRVRAATAARNHRADG 91

Search completed: June 28, 2001, 11:55:38  
Job time: 159 sec



\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:54:14 ; Search time 66.55 Seconds  
(without alignments)  
14.575 Million cell updates/sec

Title: US-09-439-313-562  
Perfect score: 88  
Sequence: 1 DWDTSALAPYLGTEEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	100.0	553	19	AAW71869
2	88	100.0	553	19	AAW69385
3	88	100.0	553	21	AAW28527
4	88	100.0	553	21	AAW82002
5	46	52.3	4473	17	AAW97244
6	42	47.7	22	16	AAW85685
7	42	47.7	22	17	AAW92886
8	42	47.7	22	18	AAW18187
9	42	47.7	22	19	AAW75575
10	42	47.7	121	19	AAW58854
11	42	47.7	527	19	AAW54152

RESULT	1
AAW71869	
ID	AAW71869 standard; Protein; 553 AA.
XX	AAW71869;
AC	AAW71869;
XX	
DT	06-JAN-1999 (first entry)
XX	
DE	Amino acid encoded by prostate tumour clone L1-12.
XX	
KW	Prostate; cancer; tumour; vaccine; immunogen; clone.
XX	
OS	Homo sapiens.
XX	

ALIGNMENTS

12	41	46.6	53	22	AAW87377	Human gene 36 enco
13	41	46.6	53	22	AAW87407	Human gene 36 enco
14	41	46.6	53	22	AAW87408	Human gene 36 enco
15	41	46.6	54	21	AAW28048	Human secreted pro
16	41	46.6	54	22	AAW75548	Human secreted pro
17	39	44.3	122	21	AAW41270	Human ORFX ORF1034
18	39	44.3	151	20	AAW85726	Novel protein (Clo
19	39	44.3	524	17	AAW99132	Mutant human tissu
20	39	44.3	524	17	AAW99133	Mutant human tissu
21	39	44.3	524	17	AAW99134	Mutant human tissu
22	39	44.3	524	17	AAW99136	Mutant human tissu
23	39	44.3	527	17	AAW99124	Mutant human tissu
24	39	44.3	527	17	AAW99125	Mutant human tissu
25	39	44.3	527	17	AAW99126	Mutant human tissu
26	39	44.3	527	17	AAW99128	Mutant human tissu
27	39	44.3	527	19	AAW54150	t-PA mutant (S127C
28	39	44.3	558	12	AAW13153	t-PA with -ve char
29	39	44.3	559	12	AAW13151	T-PA with -ve char
30	39	44.3	561	12	AAW12367	T-PA with -ve char
31	39	44.3	562	12	AAW12343	T-PA with -ve char
32	39	44.3	734	13	AAW20210	Diacylglyceroal ki
33	38	43.2	65	21	AAW00538	Human secreted pro
34	38	43.2	87	21	AAW08408	Amino acid sequenc
35	38	43.2	118	18	AAW13516	Anti-melanoma anti
36	38	43.2	209	17	AAW04249	4070A retrovirus a
37	38	43.2	209	17	AAW04250	10A1 murine leukae
38	38	43.2	288	22	AAW19865	Amphotropic MLV su
39	38	43.2	293	8	AAW70183	Sequence encoded b
40	38	43.2	310	9	AAW80770	Recombinant human
41	38	43.2	314	10	AAW90174	Tissue plasminogen
42	38	43.2	356	18	AAW28537	mt-PA9. Homo sapi
43	38	43.2	356	18	AAW14432	Tissue plasminogen
44	38	43.2	364	22	AAW19868	Activating polypep
45	38	43.2	409	16	AAW67652	Glucomylase. Sacc
46	38	43.2	433	9	AAW81914	Chain 1 of modifie
47	38	43.2	437	10	AAW94413	Sequence of coding
48	38	43.2	437	10	AAW94412	Sequence of coding
49	38	43.2	437	10	AAW94410	Sequence of coding
50	38	43.2	438	10	AAW94409	Sequence of coding
51	38	43.2	439	16	AAW68851	Delta 2-89 tissue
52	38	43.2	439	19	AAW52816	Tissue plasminogen
53	38	43.2	445	11	AAW09285	Sequence of tissue
54	38	43.2	446	11	AAW08150	Non-glycosylated t
55	38	43.2	458	18	AAW26323	Murine leukaemia v
56	38	43.2	464	13	AAW20020	t-PA Delta2-89 PA-e
57	38	43.2	472	10	AAW94416	Sequence of coding
58	38	43.2	472	19	AAW75721	Fibroblast growth
59	38	43.2	472	20	AAW49694	Human FGFR1-tpa fu
60	38	43.2	472	20	AAW06393	Human FGFR1-tpa fu
61	38	43.2	472	20	AAW06308	FGFR1-tpa fusion p
62	38	43.2	472	21	AAW12449	Human FGFR1-tpa fu
63	38	43.2	472	21	AAW32343	Human FGFR1-tpa fu
64	38	43.2	476	11	AAW03004	Tissue plasminogen
65	38	43.2	479	9	AAW81036	Modified tissue pl

Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers

Example 1; Page 87-89; 141pp; English.

This sequence is encoded by a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as this protein sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers.

Sequence 553 AA;

Query Match 100.0%; Score 88; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 8.6e-07; Indels 0;  
Matches 16; Conservative 0; Mismatches 0; Gaps 0;

QY 1 DWDTSALAPYLGTQEE 16  
Db 182 dwdtsalapylgtqee 197

RESULT 3  
AAB28527  
ID AAB28527 standard; Protein; 553 AA.  
XX  
AC AAB28527;  
XX  
DT 07-FEB-2001 (first entry)  
DE Protein encoded by human breast tumour cDNA clone P501S.  
XX  
KW Human; breast tumour antigen; cytostatic; immunotherapy;  
KW breast cancer; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200061756-A2.  
PD 19-OCT-2000.  
PF 10-APR-2000; 2000WO-US09688.  
XX  
PR 09-APR-1999; 99US-0288950.  
PR 02-JUL-1999; 99US-0346327.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Reed SG, Xu J, Dillon DC;  
XX  
OS WPI; 2000-638568/61.  
DR N-PSDB; AAC79473.  
XX  
PT A novel isolated polypeptide comprising an immunogenic portion of a breast cancer protein useful in the detection and treatment of breast cancer.  
XX  
PS Claim 2; Page 92-93; 95pp; English.  
XX  
PS The present sequence is encoded by a cDNA sequence which was isolated from a breast tumour cDNA library. It is provided in a specification relating to compounds for immunotherapy and diagnosis of breast cancer.  
CC  
CC Breast tumour antigens and the polynucleotides that encode them may be used in the production of a pharmaceutical composition to be used in the treatment of breast cancer. Proliferated T cells and incubated antigen presenting cells are also required. The polypeptides and polynucleotides may also be used to produce a vaccine.  
CC  
XX

PN WO9837093-A2.  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03492.  
XX  
PR 09-FEB-1998; 98US-0020956.  
PR 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
DR WPI; 1998-609886/51.  
DR N-PSDB; AAV61201.  
XX  
PT Polypeptides comprising immunogenic portions of prostate proteins used in a vaccine for the treatment of prostate cancer  
PT  
XX  
PS Example 1; Page 82-84; 130pp; English.  
XX  
CC The present sequence is an immunogenic portion of a prostate tumour protein. The immunogen, or the DNA encoding it, can be used as a vaccine for the treatment of prostate cancer. The immunogen was isolated from a prostate tumour cDNA library obtained by subtracting a prostate tumour cDNA expression library with a normal tissue cDNA library.  
CC  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 88; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 8.6e-07; Indels 0;  
Matches 16; Conservative 0; Mismatches 0; Gaps 0;

QY 1 DWDTSALAPYLGTQEE 16  
Db 182 dwdtsalapylgtqee 197

RESULT 2  
AAW69385  
ID AAW69385 standard; Protein; 553 AA.  
XX  
AC AAW69385;  
XX  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone L1-12 protein.  
XX  
KW Prostate tumour specific gene; human; prostate cancer; detection; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO9837418-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; 98WO-US03690.  
XX  
PR 09-FEB-1998; 98US-0904809.  
PR 25-FEB-1997; 97US-0806596.  
PR 01-AUG-1997; 97US-0904809.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
DR WPI; 1998-480805/41.  
DR N-PSDB; AAV58586.  
XX

SQ Sequence 553 AA;

Query Match 100.0%; Score 88; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 8.6e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DWDTSalAPYLGTQEE 16  
|||||  
Db 182 dwdtsalapylgtqee 197

RESULT 4  
AAY82002  
ID AAY82002 standard; Protein; 553 AA.  
XX  
AC AAY82002;  
XX  
DT 13-JUN-2000 (first entry)  
XX  
DE Human immunogenic prostate tumour protein sequence SEQ ID NO:113.  
XX  
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
KW immunogenic; cytostatic; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200004149-A2.  
XX  
PD 27-JAN-2000.  
XX  
PF 14-JUL-1999; 99WO-US15838.  
XX  
PR 14-JUL-1998; 98US-0115453.  
PR 14-JUL-1998; 98US-0116134.  
PR 23-SEP-1998; 98US-0159812.  
PR 23-SEP-1998; 98US-0159822.  
PR 15-JAN-1999; 99US-0232149.  
PR 15-JAN-1999; 99US-0232880.  
PR 09-APR-1999; 99US-0288946.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;  
XX  
DR WPI; 2000-171268/15.  
XX  
PT New polypeptide useful for treating and diagnosing prostate cancer  
PT comprises an immunogenic portion of prostate tumor protein -  
XX  
PS Claim 3; Page 138-139; 263pp; English.  
XX  
CC The present invention describes isolated polypeptides, comprising an  
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides  
CC and polynucleotides encoding them have cytostatic activity and can be  
CC used in vaccines and in gene therapy. The polypeptides and  
CC polynucleotides encoding them, antigen presenting cells which express  
CC the polypeptides, antibodies against the polypeptides and vaccines  
CC comprising them can be used for inhibiting the development of prostate  
CC cancer in a patient. The polypeptides can be used to generate antibodies  
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of  
CC the polynucleotides encoding the polypeptides can be used as a probe or  
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and  
CC AAY82000 to AAY82020 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 88; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 8.6e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DWDTSalAPYLGTQEE 16  
|||||  
Db 182 dwdtsalapylgtqee 197

RESULT 5  
AAR97244  
ID AAR97244 standard; Protein; 4473 AA.  
XX  
AC AAR97244;  
XX  
DT 07-JAN-1997 (first entry)  
XX  
DE Virulence gene cluster polypeptide product.  
XX  
KW Mutant; adaptation; virulence factor; identification; screening;  
KW vaccine; drugs; infection; treatment.  
XX  
OS Salmonella typhimurium.  
XX  
FH Key Location/Qualifiers  
FT Region /note= "All x's in this sequence correspond to  
FT termination codons in the virulence gene  
FT cluster sequence given in AAT09224."  
XX  
PN WO9617951-A2.  
XX  
PD 13-JUN-1996.  
XX  
PF 11-DEC-1995; 95WO-GB02875.  
XX  
PR 05-MAY-1995; 95GB-0009239.  
PR 09-DEC-1994; 94GB-0024921.  
PR 31-JAN-1995; 95GB-0001881.  
XX  
PA (RPMS-) RPMS TECHNOLOGY LTD.  
XX  
PI Holden DW;  
XX  
DR WPI; 1996-287194/29.  
DR N-PSDB; AAT09224.  
XX  
PT Identifying virulence genes in microorganisms - by introducing  
PT mutants with insertion inactivated genes into environment and  
PT retrieval and analysis of mutants  
XX  
PS Claim 51; Figure 11; 131pp; English.  
XX  
CC A method for identifying a microorganism having a reduced adaptation  
CC to a particular environment comprising the steps of: (1) providing a  
CC plurality of microorganisms each of which is independently mutated by  
CC the insertional inactivation of a gene with a nucleic acid comprising  
CC a unique marker sequence so that each mutant contains a different  
CC marker sequence, or clones of the said microorganism; (2) providing  
CC individually a stored sample of each mutant produced by step (1) and  
CC providing individually stored nucleic acid comprising the unique  
CC marker sequence from each individual mutant; (3) introducing a  
CC plurality of mutants produced by step (1) into the said particular  
CC environment and allowing those microorganisms which are able to do so  
CC to grow in the said environment; (4) retrieving microorganisms from  
CC the said environment or a selected part thereof and isolating the  
CC nucleic acid from the retrieved microorganisms; (5) comparing any  
CC marker sequences in the nucleic acid isolated in step (4) to the  
CC unique marker sequence of each individual mutant stored as in step  
CC (2); and (6) selecting an individual mutant which does not contain any  
CC of the marker sequences as isolated in step (4). The products and  
CC methods can be used for identifying virulence genes in microorganisms.  
CC The mutant microorganisms can be used in vaccines or to screen for  
CC drugs which reduce virulence or compounds useful for preventing,  
CC ameliorating or treating infections in animals or plants.  
XX  
SQ Sequence 4473 AA;



Query Match 52.3%; Score 46; DB 17; Length 4473;  
Best Local Similarity 70.0%; Pred. NO. 1.2e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 DWDTSALAPY 10  
Db 2979 ewdskalap 2988

RESULT 6  
AAR85685  
ID AAR85685 standard; peptide; 22 AA.

XX AC AAR85685;

XX DT 25-JAN-1996 (first entry)

XX DE 23 kD M.tuberculosis extracellular protein N-terminal fragment.

XX XX Majorly abundant extracellular product; MAP; M.tuberculosis;  
KW immunise; vaccinating agent; vaccine; M.bovis; M.marinum; M.kansasii;  
KW M.aviumintracellulare; M.fortuitum; M.chelonae; M.scrofulaceum;  
KW M.leprae; M.africanum; M.ulcerans; M.microti; human; cat; dog;  
KW cattle; sheep; horse; pig.

XX OS Mycobacterium tuberculosis.

XX PN WO9514713-A2.

XX PD 01-JUN-1995.

XX PF 18-NOV-1994; 94WO-US13145.

XX PR 12-AUG-1994; 94US-0289667.

XX PT 23-NOV-1993; 93US-0156358.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Horwitz MA;

XX PS WPI; 1995-206898/27.

XX XX Vaccine composed of major extracellular prods. of Mycobacterium  
PT used to promote immune response to various Mycobacterium pathogens,  
PT such as M.tuberculosis, M. bovis, M. leprae, etc. in humans,  
PT cattle, sheep, etc.

XX PS Claim 12; Page 103; 123pp; English.

XX XX The sequences given in AAR85680-86 represent N-terminal peptides from  
CC majorly abundant extracellular products (MAPs) produced by M.  
CC tuberculosis which were used in the vaccinating agent of the  
CC invention. MAPs constitute approx. 90% of all of the proteins  
CC released by M.tuberculosis in culture. MAPs can be used to immunise  
CC a host animal, and due to their release from M.tuberculosis when it  
CC infects the host, they are presented to the host immune system at high  
CC frequency. Due to their profuse and continual presentation to the  
CC infected host's immune system, the most prevalent bacterial  
CC extracellular products provoke a vigorous immune response largely  
CC irrespective of their individual molecular immunogenic characteristics.  
CC The vaccinating agent may be used to provide immunity against M.  
CC tuberculosis, M.bovis, M.marinum, M.kansasii, M.aviumintracellulare,  
CC M.fortuitum, M.chelonae, M.scrofulaceum, M.leprae, M.africanum, M.  
CC ulcerans, and M.microti in humans, cats, dogs, cattle, sheep, horses  
CC and pigs.

XX XX Sequence 22 AA;

Query Match 47.7%; Score 42; DB 16; Length 22;  
Best Local Similarity 50.0%; Pred. NO. 1.6;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
OY 1 DWDTSALAPYLGTO 14  
Db 9 dwdygalephisgq 22

RESULT 7  
AAR92886  
ID AAR92886 standard; peptide; 22 AA.

XX AC AAR92886;

XX DT 30-SEP-1996 (first entry)

XX DE Mycobacterium 23 kD protein N-terminal peptide.

XX XX N-terminal peptide; extracellular product; Mycobacterium; 23 kD protein;  
KW immunoreactive; vaccine; immune response; pathogen; mammal; therapy;  
KW opsonising humoral response; intracellular bacteria.

XX OS Mycobacterium tuberculosis.

XX PN WO9605223-A1.

XX PD 22-FEB-1996.

XX PF 24-FEB-1995; 95WO-US02373.

XX PR 12-AUG-1994; 94US-0289667.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Horwitz MA;

XX PS WPI; 1996-139640/14.

XX XX Vaccines comprising abundant extracellular prods. - useful to  
PT vaccinate against intracellular bacteria, partic. Mycobacterium  
PT tuberculosis.

XX PS Claim 59; Page 109; 138pp; English.

XX XX AAR92876-R92889 represent N-terminal fragments of the abundant  
CC extracellular products of Mycobacterium. This sequence represents the  
CC N-terminus of the 23 kD protein of M. tuberculosis. The 23 kD protein,  
CC or an immunoreactive homologue or a fragment of it, with this sequence  
CC as the N-terminus can be used in a vaccinating agent of the invention.  
CC The vaccinating agent promotes an immune response against a  
CC Mycobacterium pathogen, and are therefore useful for immunising a mammal  
CC against one of these pathogens. As the vaccines do not contain whole  
CC bacteria or components, they are less likely to provoke toxic responses  
CC compared to attenuated or killed bacterial vaccines. Also, the vaccines  
CC are not life-threatening to immunocompromised individuals. The use of  
CC extracellular products also prevents the development of an opsonising  
CC humoral response which can increase the pathogenesis of intracellular  
CC bacteria. It also reduces the potential for generating a response which  
CC precludes the use of widely used screening and control techniques based  
CC on host recognition of immunogenic agents.

XX XX Sequence 22 AA;

Query Match 47.7%; Score 42; DB 17; Length 22;  
Best Local Similarity 50.0%; Pred. NO. 1.6;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLGTO 14  
Db 9 dwdygalephisgq 22

RESULT 8

AAW18187  
ID AAW18187 standard; peptide; 22 AA.

XX AAW18187;

DT 13-AUG-1997 (first entry)

DE N-terminal sequence from Mycobacterium tuberculosis 23KD protein.

XX Vaccine; vaccinating agent; M.tuberculosis; pathogen; bacteria;  
KW virus; fungus; protozoan; HIV.

OS Mycobacterium tuberculosis.

XX WO9637219-A1.

PN 28-NOV-1996.

PD 23-MAY-1996; 96WO-US07781.

XX 06-DEC-1995; 95US-0568357.

PR 23-MAY-1995; 95US-0447398.

PR 20-OCT-1995; 95US-0545926.

PR 31-OCT-1995; 95US-0551149.

XX (REGC ) UNIV CALIFORNIA.

PA Harth G, Horwitz MA;

XX WPI; 1997-020936/02.

XX Vaccines derived from M.tuberculosis major abundant extracellular

PS Example 2; Page 110; 193pp; English.

XX A vaccinating agent for promoting an immune response in a mammal  
CC against Mycobacterium pathogens comprises at least one  
CC immunodominant epitope of at least one majorly abundant  
CC extracellular protein, i.e. the M.tuberculosis 110, 80, 71, 58, 45,  
CC 32A, 32B, 30, 24, 23.5, 23, 16, 14 or 12 kD proteins, or their  
CC analogues, homologues and subunits. The present sequence represents  
CC the N-terminal amino acid sequence of the 23 kD protein. The N-terminal  
CC amino acid sequences of the major abundant extracellular products were  
CC determined to provide structural data and to uncover possible  
CC relationships between the proteins. The vaccinating agents are used to  
CC protect against (or to treat existing) infections by Mycobacterium  
CC (especially M.tuberculosis) while the epitopes can also be used to  
CC detect presence of an immune response to a Mycobacterium pathogen. The  
CC vectors, containing the DNA for the extracellular proteins, are used to  
CC transform cells for production of recombinant DNA molecules. More  
CC generally the DNA from other pathogens can be used in vaccines, e.g.  
CC against other bacteria, viruses, fungi and protozoa. Since different  
CC combinations of DNA can be used, a wide range of effective compositions  
CC can be produced. They generate a response against the antigens most  
CC often found on infected cells during the infection, regardless of the  
CC strength or specificity of the immune response. The vaccines are easy to  
CC produce and less toxic than known killed or attenuated vaccines, so can  
CC be given to immunocompromised subjects, e.g. those with HIV infection.

XX Sequence 22 AA;

Query Match 47.7%; Score 42; DB 18; Length 22;  
Best Local Similarity 50.0%; Pred. NO. 1.6;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTO 14

Db 9 dwdygalephisgq 22

RESULT 9

AAW75575

ID AAW75575 standard; protein; 22 AA.

XX AAW75575;

DT 23-OCT-1998 (first entry)

DE M. tuberculosis 23 kD extracellular product N-terminal sequence.

XX Mycobacterium tuberculosis; vaccination; extracellular product;  
KW immunodominant epitope; interleukin-12; MF59; immune response;  
KW opsonising humoral response; intracellular pathogen.

XX Mycobacterium tuberculosis.

PN WO9831388-A1.

PD 23-JUL-1998.

PF 15-JAN-1998; 98WO-US00942.

PR 21-JAN-1997; 97US-0786533.

XX (REGC ) UNIV CALIFORNIA.

PA Harth G, Horwitz MA, Lee B;

XX WPI; 1998-413815/35.

XX Vaccines against Mycobacterium containing major extracellular  
PT proteins - used to, e.g. induce protective and therapeutic immune  
PT responses, and for detecting an immune response

PS Example 2; Page 35; 236pp; English.

XX Sequences shown in AAW75570 to AAW75586 represent N-terminal amino acid  
CC sequences of 14 exemplary major abundant extracellular products of  
CC Mycobacterium tuberculosis. The invention provides an agent for  
CC vaccinating mammals against Mycobacterium. The agent comprises at least  
CC one of the major abundant extracellular 110, 80, 71, 58, 45, 32A, 32B,  
CC 30, 24, 23.5, 23, 16, 14 or 12 kDa proteins of M. tuberculosis, or at  
CC least 1 of their immunodominant epitopes and interleukin-12 (IL-12) or  
CC MF59 as adjuvants. The agent containing the nucleic acid encoding the  
CC extracellular products are used to raise a protective or therapeutic  
CC immune response against Mycobacterium, specifically M. tuberculosis.  
CC The immunodominant epitopes can also be used (typically in a cutaneous  
CC hypersensitivity test) to detect an immune response to vaccination.  
CC Preparation of the agent does not require selection of the most  
CC immunogenic products, so large scale production and purification are  
CC easy, resulting in a consistent, standardised formulation, having lower  
CC toxicity than killed or attenuated vaccines. The agents provide a rapid  
CC and effective response (including a strong cell-mediated component) and  
CC are safe even in immunocompromised subjects. They prevent development of  
CC an opsonising humoral response that might spread intracellular  
CC pathogens.

XX Sequence 22 AA;

Query Match 47.7%; Score 42; DB 19; Length 22;  
Best Local Similarity 50.0%; Pred. No. 1.6;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTO 14

Db 9 dwdygalephisgq 22

RESULT 10

AAW58854

ID AAW58854 standard; Protein; 121 AA.

[illegible]

DE Human gene 36 encoded secreted protein HLHDL42, SEQ ID NO:118.  
XX  
KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulneryary;  
KW cell culture; chemotaxis; food additive;  
KW binding partner identification.  
XX  
OS Homo sapiens.  
XX  
XX WO200118022-A1.  
PN  
XX  
PD 15-MAR-2001.  
XX  
XX 31-AUG-2000; 2000WO-US24008.  
PF  
XX 03-SEP-1999; 99US-0152315.  
PR  
PR 03-SEP-1999; 99US-0152317.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;  
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;  
PI Moore PA, Shi Y, Wei Y, Florence KA;  
XX  
DR WPI; 2001-203081/20.  
DR N-PSDB; AAF91893.  
XX  
XX Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -  
XX  
PS Claim 11; Page 553; 607pp; English.  
XX  
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted  
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.  
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes  
CC and their corresponding secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 52 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein of the invention.  
XX  
SQ Sequence 53 AA;

Query Match 46.6%; Score 41; DB 22; Length 53;  
Best Local Similarity 72.7%; Pred. No. 6.1;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 5 SALAPYLGQTQE 15  
Db 41 salqpylgkee 51  
||| |||| :|  
RESULT 13  
AAB87407  
ID AAB87407 standard; Protein; 53 AA.  
XX  
AC AAB87407;  
XX  
DT 22-MAY-2001 (first entry)  
XX  
DE Human gene 36 encoded secreted protein HAPQU71, SEQ ID NO:148.  
XX  
KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulneryary;  
KW cell culture; chemotaxis; food additive;  
KW binding partner identification.  
XX  
OS Homo sapiens.  
XX  
XX WO200118022-A1.  
PN  
XX  
PD 15-MAR-2001.  
XX  
PF 31-AUG-2000; 2000WO-US24008.  
XX  
PR 03-SEP-1999; 99US-0152315.  
PR 03-SEP-1999; 99US-0152317.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;  
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;  
PI Moore PA, Shi Y, Wei Y, Florence KA;  
XX  
XX WPI; 2001-203081/20.  
DR N-PSDB; AAF91923.  
XX  
XX Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -  
XX  
PS Claim 11; Page 574; 607pp; English.  
XX  
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted  
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.  
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes  
CC and their corresponding secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 52 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein of the invention.  
XX  
SQ Sequence 53 AA;



CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein of the invention.

XX Sequence 53 AA;  
SQ

Query Match 46.6%; Score 41; DB 22; Length 53;  
Best Local Similarity 72.7%; Pred. No. 6.1;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 SALAPYLGTQE 15  
Db 41 salqpylgkee 51  
||| |||| :|

RESULT 14  
AAB87408  
ID AAB87408 standard; Protein; 53 AA.  
XX  
AC AAB87408;  
XX  
XX 22-MAY-2001 (first entry)  
XX Human gene 36 encoded secreted protein HAPQU71, SEQ ID NO:149.  
DE  
DE Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulneryary;  
KW cell culture; chemotaxis; food additive;  
KW binding partner identification.  
XX  
XX Homo sapiens.  
XX  
XX WO200118022-A1.  
PN  
XX  
XX 15-MAR-2001.  
XX  
XX 31-AUG-2000; 2000WO-US24008.  
PF  
XX  
XX 03-SEP-1999; 99US-0152315.  
PR  
XX 03-SEP-1999; 99US-0152317.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;  
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;  
PI Moore PA, Shi Y, Wei Y, Florence KA;  
XX  
XX WPI: 2001-203081/20.  
DR N-PSDB; AAF91924.  
DR  
XX Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -  
XX  
XX Claim 11; Page 574; 607pp; English.

XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted  
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.  
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes  
CC and their corresponding secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 52 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, diabetes, atherosclerosis,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, kidney disorders,  
CC cardiovascular disorders, angiogenic disorders, pregnancy-related disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein of the invention.

XX Sequence 53 AA;  
SQ

Query Match 46.6%; Score 41; DB 22; Length 53;  
Best Local Similarity 72.7%; Pred. No. 6.1;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 SALAPYLGTQE 15  
Db 41 salqpylgkee 51  
||| |||| :|

RESULT 15  
AAB28048  
ID AAB28048 standard; protein; 54 AA.  
XX  
XX AAB28048;  
AC  
XX  
XX 02-FEB-2001 (first entry)  
XX Human secreted protein SEQ ID NO: 96.  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulneryary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX  
XX Homo sapiens.  
XX  
XX WO200055177-A2.  
PN  
XX  
XX 21-SEP-2000.  
PD  
XX  
XX 09-MAR-2000; 2000WO-US06058.  
PF  
XX  
XX 12-MAR-1999; 99US-0124145.  
PR  
XX 03-DEC-1999; 99US-0168654.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Rosen CA, Ruben SM, Komatsoulis G;  
PI

XX WPI; 2000-638177/61.  
DR N-PSDB; AAC59144.  
DR  
XX  
PT Novel nucleic acids encoding 49 human secreted proteins useful for  
PT treating cancers, hyperproliferative disorders, inflammatory disorders,  
PT neurological disorders and cardiovascular disorders  
XX  
PS Claim 11; Page 359; 389pp; English.  
XX  
CC Sequences AAB28012-B28060 represent the amino acid sequences of 49  
CC human secreted proteins encoded by the genes AAC59108-C59156. The genes  
CC and proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
CC a range of human tissues disclosed in the specification. The nucleic  
CC acids, proteins, antibodies and (ant)agonists are useful in the  
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
XX  
SQ Sequence 54 AA;

Query Match 46.6%; Score 41; DB 21; Length 54;  
Best Local Similarity 72.7%; Pred. No. 6.2;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 SALAPYLGTOE 15  
||| |||| :|  
Db 41 salqpylgkee 51

RESULT 16  
AAB75548  
ID AAB75548 standard; Protein; 54 AA.  
XX  
AC AAB75548;

XX 06-APR-2001 (first entry)

DE Human secreted protein sequence encoded by gene 43 SEQ ID NO:102.

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neutropic; neuroprotective; antibacterial; virucide; fungicide;  
KW opthalmallogical; vulnerary; autoimmune disease; cardiovascular disorder;  
KW hyperproliferative disorders; cerebrovascular disorder; wound healing;  
KW nervous system disorder; ocular disorder; skin ageing; chemotaxis;  
KW food additive.

XX Homo sapiens.

OS WO200077026-A1.

PN 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US14973.

XX 11-JUN-1999; 99US-0138630.

PR (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsoulis GA;

PI WPI; 2001-071258/08.

DR

DR N-PSDB; AAF64218.  
XX  
PT Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT parkinson's diseases and cancers  
XX  
PS Claim 11; Page 489; 542pp; English.  
XX  
CC Human secreted proteins AAB75506 - AAB75554 are encoded by polynucleotide  
CC sequences AAF64176 - AAF64224. The specification includes amino acid  
CC sequences AAB75555 - AAB75606 which represent fragments of the human  
CC secreted proteins, and protein sequences with which they share homology.  
CC The proteins and polynucleotides, their agonists and antagonists have  
CC activities dependent on the tissues and cells in which they are  
CC expressed, examples of these activities include, immunosuppressive;  
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;  
CC vasotropic; cerebroprotective; neutropic; neuroprotective; antibacterial;  
CC virucide; fungicide; opthalmallogical; and vulnerary. The proteins,  
CC polynucleotides, agonists and antagonists can be used to treat or detect  
CC or diagnose various diseases and disorders including, autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders  
CC e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,  
CC infections caused by bacteria, viruses and fungi and ocular disorders  
CC e.g. corneal infection. The polypeptides can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin ageing due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. Included in the invention are  
CC polynucleotide sequences AAF64167 - AAF64175 and peptide AAB75505 which  
CC are used in the isolation, identification and characterisation of the  
CC proteins of the invention.  
XX  
SQ Sequence 54 AA;

Query Match 46.6%; Score 41; DB 22; Length 54;  
Best Local Similarity 72.7%; Pred. No. 6.2;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 SALAPYLGTOE 15  
||| |||| :|  
Db 41 salqpylgkee 51

RESULT 17  
AAB41270

ID AAB41270 standard; Protein; 122 AA.

XX AAB41270;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1034 polypeptide sequence SEQ ID NO:2068.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnerary; antipsoriatic; antiparkinsonian; neutropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.

XX Homo sapiens.

OS

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U93236; AAC51228.1; -.  
DR EMBL; U93237; AAC51229.1; -.  
DR EMBL; U93237; AAC51230.1; -.  
DR MIM; I31100; -.  
DR MIM; I45000; -.  
KW Nuclear protein; Disease mutation; Alternative splicing; Polymorphism.  
FT VARSPLIC 149 153 MISSING (IN SHORT ISOFORM).  
FT VARIANT 12 12 P -> L (IN FMEN1).  
FT VARIANT 22 22 /FTId=VAR\_005425.  
FT VARIANT 22 22 L -> R (IN FMEN1).  
FT VARIANT 26 26 /FTId=VAR\_005426.  
FT VARIANT 26 26 E -> K (IN PARATHYROID ADENOMA).  
FT VARIANT 39 39 /FTId=VAR\_005427.  
FT VARIANT 39 39 L -> W (IN FMEN1).  
FT VARIANT 42 42 /FTId=VAR\_005428.  
FT VARIANT 42 42 G -> D (IN FMEN1).  
FT VARIANT 45 45 /FTId=VAR\_005429.  
FT VARIANT 45 45 E -> G (IN FMEN1).  
FT VARIANT 119 119 /FTId=VAR\_005430.  
FT VARIANT 135 135 MISSING (IN FMEN1).  
FT VARIANT 135 135 /FTId=VAR\_005431.  
FT VARIANT 139 139 K -> I (IN FMEN1).  
FT VARIANT 139 139 /FTId=VAR\_005434.  
FT VARIANT 139 139 H -> D (IN FMEN1).  
FT VARIANT 139 139 /FTId=VAR\_005432.  
FT VARIANT 139 139 H -> Y (IN FMEN1 AND SPORADIC MEN1).  
FT VARIANT 144 144 /FTId=VAR\_005433.  
FT VARIANT 144 144 F -> V (IN FMEN1).  
FT VARIANT 161 161 /FTId=VAR\_005436.  
FT VARIANT 161 161 G -> D (IN FMEN1).  
FT VARIANT 165 165 /FTId=VAR\_008017.  
FT VARIANT 165 165 A -> P (IN FMEN1).  
FT VARIANT 169 169 /FTId=VAR\_005437.  
FT VARIANT 171 173 /FTId=VAR\_005438.  
FT VARIANT 176 176 MISSING (IN FMEN1).  
FT VARIANT 176 176 /FTId=VAR\_005439.  
FT R -> Q.

Query Match 31.5%; Score 46; DB 1; Length 615;  
Best Local Similarity 52.9%; Pred. No. 47;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEA 18  
Db 512 GQGAVSGPPRKPGTVA 528

RESULT 13  
ID P73\_CERAE STANDARD; PRT; 637 AA.  
AC Q9XSK8; Q9TSQ9;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE TUMOR PROTEIN P73 (P53-LIKE TRANSCRIPTION FACTOR) (P53-RELATED  
DE PROTEIN).  
GN TP73 OR P73.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
OC Cercopithecinae; Cercopithecus.  
OX NCBI\_TaxID=9534;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;

RA Caput D.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PARTICIPATES IN THE APOPTOTIC RESPONSE TO DNA DAMAGE.  
CC WHEN OVERPRODUCED, ACTIVATES TRANSCRIPTION FROM P53-RESPONSIVE  
CC PROMOTERS AND INDUCES APOPTOSIS. MAY BE A TUMOR SUPPRESSOR  
CC PROTEIN (BY SIMILARITY).  
CC -!- SUBUNIT: THE C-TERMINAL OLIGOMERIZATION DOMAIN BINDS TO THE ABL  
CC TYROSINE KINASE SH3 DOMAIN. ISOFORM BETA INTERACTS HOMOTYPICALLY  
CC AND WITH P53, WHEREAS ISOFORM ALPHA DOES NOT.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- DOMAIN: POSSESSES AN ACIDIC TRANSACTIVATION DOMAIN, A CENTRAL DNA  
CC BINDING DOMAIN AND A C-TERMINAL OLIGOMERIZATION DOMAIN THAT BINDS  
CC TO THE ABL TYROSINE KINASE SH3 DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y11419; CAA72224.1; -.  
DR EMBL; Y11419; CAA72225.1; -.  
DR HSSP; P04637; 1YCS.  
DR InterPro; IPR002117; -.  
DR Pfam; PF00870; P53; 1.  
DR PRINTS; PR00386; P53SUPPRESSR.  
KW Transcription regulation; Activator; DNA-binding; Anti-oncogene;  
KW Apoptosis; Nuclear protein; Phosphorylation; Alternative splicing.  
FT VARSPLIC 495 637  
FT SFLTGLGCPNCIEYFTSQGLQSIYHLQNLTIEDLGALKIPE  
FT QYRMTIWRGLQDLKQHDYGAQAQQLRSSNAAAIISIGSG  
FT ELQQRVMEAVHFRVVRHTITIPNRGGPGAGPDEWADFGFDL  
FT PDCRKARKQPIKEEFTAEIH -> RTWGP (IN ISOFORM  
FT BETA).  
SQ SEQUENCE 637 AA; 69630 MW; 7CB200B919C9C70A CRC64;

Query Match 31.5%; Score 46; DB 1; Length 637;  
Best Local Similarity 56.2%; Pred. No. 48;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 9 VPRAEPGTEARRHYDE 24  
Db 337 VPALGPGVKRRHGD 352

RESULT 14  
ID IE63\_HCMVA STANDARD; PRT; 744 AA.  
AC P16749;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE TRANSCRIPTIONAL REGULATOR IE63 HOMOLOG (PROTEIN UL69).  
GN UL69.  
OS Human cytomegalovirus (strain AD169).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90269039; PubMed=2161319;  
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
RA Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,  
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
RT "Analysis of the protein-coding content of the sequence of human  
RT cytomegalovirus strain AD169.";  
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).  
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,

FT Misc-difference 117 /label= substitution  
FT /note= "Asn to Glu substitution"  
FT Misc-difference 184  
FT /label= substitution  
FT /note= "Asn to Glu substitution"  
FT Misc-difference 277  
FT /label= substitution  
FT /note= "Lys to Arg substitution"  
FT Misc-difference 448  
FT /label= substitution  
FT /note= "Asn to Glu substitution"  
FT Misc-difference 503..524  
FT /label= deletion  
FT /note= "amino acids 503 to 524 are opt. deleted"

XX  
PN US5501853-A.  
XX  
PD 26-MAR-1996.

XX  
PF 23-DEC-1985; 85US-0812879.  
XX  
PR 22-DEC-1986; 86US-0944117.  
PR 23-DEC-1985; 85US-0812879.

XX  
PA (BEHW ) BEHRINGWERKE AG.  
PA (CHIR ) CHIRON CORP.

XX  
PI Afting E, Haigwood NL, Mullenbach G, Paques EP;  
XX WPI; 1996-179247/18.

XX  
PT Human tissue plasminogen activator mutants for fibrin clot lysis -  
PT with improved plasminogenolytic specific activity, increased fibrin  
PT dependence and decreased plasminogen inhibitor susceptibility

XX  
PS Claim 4; Page -: 8pp; English.

XX  
CC AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants  
CC that have improved plasminogen specific activity, increased fibrin  
CC dependence and decreased plasminogen inhibitor susceptibility. The  
CC mutants are useful in the lysis of fibrin clots and in the  
CC prevention of blood clot formation by activating plasminogen. The  
CC specific activity of tPA is enhanced by reducing the amount of  
CC glycosylation of the protein by modifying the glycosylation sites at  
CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to  
CC plasminogen activator inhibition is reduced by modifying the cleavage  
CC site at position 274-278 and the fibrin dependence of tPA is increased  
CC by truncating the C-terminus.

XX  
SQ Sequence 524 AA;

Query Match 44.3%; Score 39; DB 17; Length 524;  
Best Local Similarity 44.4%; Pred. No. 1.7e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTOEE 16  
Db :|::||| | | : :  
115 nwessalaqkypysgrpd 132

RESULT 20  
AAR99133  
ID AAR99133 standard; protein; 524 AA.

XX  
AC AAR99133;

XX 29-OCT-1996 (first entry)

DE Mutant human tissue plasminogen activator for fibrin clot lysis.

XX Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;

KW improved specificity; glycosylation sites; truncated C-terminus;  
KW deletion; substitution; fibrinogen.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 117  
FT /label= substitution  
FT /note= "Asn to Glu substitution"  
FT Misc-difference 184  
FT /label= substitution  
FT /note= "Asn to Glu substitution"  
FT Misc-difference 277  
FT /label= substitution  
FT /note= "Lys to Arg substitution"  
FT Misc-difference 503..524  
FT /label= deletion  
FT /note= "amino acids 503 to 524 are opt. deleted"

XX  
PN US5501853-A.

XX  
PD 26-MAR-1996.

XX  
PF 23-DEC-1985; 85US-0812879.

XX  
PR 22-DEC-1986; 86US-0944117.  
PR 23-DEC-1985; 85US-0812879.

XX  
PA (BEHW ) BEHRINGWERKE AG.  
PA (CHIR ) CHIRON CORP.

XX  
PI Afting E, Haigwood NL, Mullenbach G, Paques EP;  
XX WPI; 1996-179247/18.

XX  
PT Human tissue plasminogen activator mutants for fibrin clot lysis -  
PT with improved plasminogenolytic specific activity, increased fibrin  
PT dependence and decreased plasminogen inhibitor susceptibility

XX  
PS Claim 4; Page -: 8pp; English.

XX  
CC AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants  
CC that have improved plasminogen specific activity, increased fibrin  
CC dependence and decreased plasminogen inhibitor susceptibility. The  
CC mutants are useful in the lysis of fibrin clots and in the  
CC prevention of blood clot formation by activating plasminogen. The  
CC specific activity of tPA is enhanced by reducing the amount of  
CC glycosylation of the protein by modifying the glycosylation sites at  
CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to  
CC plasminogen activator inhibition is reduced by modifying the cleavage  
CC site at position 274-278 and the fibrin dependence of tPA is increased  
CC by truncating the C-terminus.

XX  
SQ Sequence 524 AA;

Query Match 44.3%; Score 39; DB 17; Length 524;  
Best Local Similarity 44.4%; Pred. No. 1.7e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTOEE 16  
Db :|::||| | | : :  
115 nwessalaqkypysgrpd 132

RESULT 21  
AAR99134  
ID AAR99134 standard; protein; 524 AA.

XX  
AC AAR99134;

XX 29-OCT-1996 (first entry)

XX



xx	29-OCT-1996	(first entry)
xx		
xx		Mutant human tissue plasminogen activator for fibrin clot lysis.
xx		
xx		
xx		Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
xx		improved specificity; glycosylation sites; truncated C-terminus;
xx		deletion; substitution; fibrinogen.

```

Synthetic.
Key
Misc-difference 117 Location/Qualifiers
                    /label= substitution
                    /note= "Asn to Glu substitution"
Misc-difference 277
                    /label= substitution
                    /note= "Lys to Arg substitution"
Misc-difference 503...524
                    /label= deletion
                    /note= "amino acids 503 to 524 a

```

US5501853-A.

26-MAR-1996.

23-DEC-1985: 85US-0812879.

22-DEC-1986: 86US-0944117.

23-DEC-1983; 8303-081207

(BEHW) BEHRINGWERKE AG  
(CHIB) CHIRON CORP

(CHIR ) CHIRON CORP.

Afting E, Haigwood NL, Mullenbach G, paques EF

WPI; 1996-179247/18.

Human tissue plasminogen activator mutants for fibrin clot lysis with improved plasminogenolytic specific activity, increased fibrin dependence and decreased plasminogen inhibitor susceptibility

Claim 4; Page -; 8pp; English.

AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants that have improved plasminogen specific activity, increased fibrin dependence and decreased plasminogen inhibitor susceptibility. The mutants are useful in the lysis of fibrin clots and in the prevention of blood clot formation by activating plasminogen. The specific activity of tPA is enhanced by reducing the amount of glycosylation of the protein by modifying the glycosylation sites at positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to plasminogen activator inhibition is reduced by modifying the cleavage site at position 274-278 and the fibrin dependence of tPA is increased by truncating the C-terminus.

Sequence 524 AA;

Query Match	44.3%	Score	NO. 1.7e+02;	Indels
Best Local	44.4%	Pred.	NO. 1.7e+02;	
Matches	8:	Conservative	5:	Mismatches
	3:			Indels

```
QY      1 DWDTSALA--PYLGTQEE 16
      :1::1111 111::
DB      115 nwessalackpvsgrpd 132
```

AAR99124

ID AAR99124 standard; protein; 527 AA.

AAR99124;

AC  
AAAR991247

```
XX 29-OCT-1996 (first entry)
DT Mutant human tissue plasminogen activator for fibrin clot lysis.
XX
DE Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
KW improved specificity; glycosylation sites; truncated C-terminus;
KW deletion; substitution; fibrinogen.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 117 /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 184
FT /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 277
FT /label= substitution
FT /note= "Lys to Arg substitution"
FT Misc-difference 448
FT /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 525
FT /label= deletion
FT /note= "Met-525 opt. deleted"
FT Misc-difference 527
FT /label= deletion
FT /note= "Pro-527 opt. deleted"
XX
PN US5501853-A.
XX
PD 26-MAR-1996.
XX
PF 23-DEC-1985; 85US-0812879.
XX
PR 22-DEC-1986; 86US-0944117.
PR 23-DEC-1985; 85US-0812879.
XX
PA (BEHW ) BEHRINGERWERKE AG.
PA (CHIR ) CHIRON CORP.
XX
PI Afting E, Haigwood NL, Mullenbach G, Paques EP;
XX
DR WPI; 1996-179247/18.
XX
PS Human tissue plasminogen activator mutants for fibrin clot lysis -
XX with improved plasminogenolytic specific activity, increased fibrin
CC dependence and decreased plasminogen inhibitor susceptibility. The
CC mutants are useful in the lysis of fibrin clots and in the
CC prevention of blood clot formation by activating plasminogen. The
CC specific activity of tPA is enhanced by reducing the amount of
CC glycosylation of the protein by modifying the glycosylation sites at
CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
CC plasminogen activator inhibition is reduced by modifying the cleavage
CC site at position 274-278 and the fibrin dependence of tPA is increased
XX by truncating the C-terminus.
SQ Sequence 527 AA;

Query Match 44.3%; Score 39; DB 17; Length 527;
Best Local Similarity 44.4%; Pred. NO. 1.7e+02;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTOEE 16
```

```
Db 115 nwessalaqkypysgrpd 132
:|::| | | | | | :
RESULT 24
AAR99125
ID AAR99125 standard; protein; 527 AA.
XX
AC AAR99125;
XX
DT 29-OCT-1996 (first entry)
XX
DE Mutant human tissue plasminogen activator for fibrin clot lysis.
XX
KW Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
KW improved specificity; glycosylation sites; truncated C-terminus;
KW deletion; substitution; fibrinogen.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 117 /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 184 /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 277 /label= substitution
FT /note= "Lys to Arg substitution"
FT Misc-difference 525 /label= deletion
FT /note= "Met-525 opt. deleted"
FT Misc-difference 527 /label= deletion
FT /note= "Pro-527 opt. deleted"
XX
PN US5501853-A.
XX
PD 26-MAR-1996.
XX
PF 23-DEC-1985; 85US-0812879.
XX
PR 22-DEC-1986; 86US-0944117.
PR 23-DEC-1985; 85US-0812879.
XX
PA (BEHW ) BEHRINGERWERKE AG.
PA (CHIR ) CHIRON CORP.
XX
PI Afting E, Haigwood NL, Mullenbach G, Paques EP;
XX
DR WPI; 1996-179247/18.
XX
PS Human tissue plasminogen activator mutants for fibrin clot lysis -
XX with improved plasminogenolytic specific activity, increased fibrin
CC dependence and decreased plasminogen inhibitor susceptibility
CC Claim 3; Page -; 8pp; English.
XX
AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
CC that have improved plasminogen specific activity, increased fibrin
CC dependence and decreased plasminogen inhibitor susceptibility. The
CC mutants are useful in the lysis of fibrin clots and in the
CC prevention of blood clot formation by activating plasminogen. The
CC specific activity of tPA is enhanced by reducing the amount of
CC glycosylation of the protein by modifying the glycosylation sites at
CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
CC plasminogen activator inhibition is reduced by modifying the cleavage
CC site at position 274-278 and the fibrin dependence of tPA is increased
XX by truncating the C-terminus.
SQ Sequence 527 AA;
```

```

Query Match      44.3%; Score 39; DB 17; Length 527;
Best Local Similarity 44.4%; Pred. No. 1.7e+02;
Matches      8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
SQ Sequence 527 AA;

QY 1 DWDTSALA--PYLGTOEE 16
Db 115 nwessalaqkpygripd 132

RESULT 25
AAR99126
ID AAR99126 standard; protein; 527 AA.
XX AAR99126;
XX 29-OCT-1996 (first entry)
XX Mutant human tissue plasminogen activator for fibrin clot lysis.
XX Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
XX improved specificity; glycosylation sites; truncated C-terminus;
XX deletion; substitution; fibrinogen.
XX Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 117
FT /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 277
FT /label= substitution
FT /note= "Lys to Arg substitution"
FT Misc-difference 448
FT /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 525
FT /label= deletion
FT /note= "Met-525 opt. deleted"
FT Misc-difference 527
FT /label= deletion
FT /note= "Pro-527 opt. deleted"
XX US5501853-A.
XX 26-MAR-1996.
XX 23-DEC-1985; 85US-0812879.
XX 22-DEC-1986; 86US-0944117.
XX 23-DEC-1985; 85US-0812879.
XX (BEHW ) BEHRINGWERKE AG.
XX (CHIR ) CHIRON CORP.
XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
XX WPI; 1996-179247/18.
XX Human tissue plasminogen activator mutants for fibrin clot lysis
XX with improved plasminogenolytic specific activity, increased fibrin
XX dependence and decreased plasminogen inhibitor susceptibility
XX Claim 3; Page -: 8pp; English.
XX AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
XX that have improved plasminogen specific activity, increased fibrin
XX dependence and decreased plasminogen inhibitor susceptibility. The
XX mutants are useful in the lysis of fibrin clots and in the
XX prevention of blood clot formation by activating plasminogen. The
XX specific activity of tPA is enhanced by reducing the amount of
XX glycosylation of the protein by modifying the glycosylation sites at
XX positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
CC

Query Match      44.3%; Score 39; DB 17; Length 527;
Best Local Similarity 44.4%; Pred. No. 1.7e+02;
Matches      8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
SQ Sequence 527 AA;

QY 1 DWDTSALA--PYLGTOEE 16
Db 115 nwessalaqkpygripd 132

RESULT 26
AAR99128
ID AAR99128 standard; protein; 527 AA.
XX AAR99128;
XX 29-OCT-1996 (first entry)
XX Mutant human tissue plasminogen activator for fibrin clot lysis.
XX Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
XX improved specificity; glycosylation sites; truncated C-terminus;
XX deletion; substitution; fibrinogen.
XX Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 117
FT /label= substitution
FT /note= "Asn to Glu substitution"
FT Misc-difference 277
FT /label= substitution
FT /note= "Lys to Arg substitution"
FT Misc-difference 525
FT /label= deletion
FT /note= "Met-525 opt. deleted"
FT Misc-difference 527
FT /label= deletion
FT /note= "Pro-527 opt. deleted"
XX US5501853-A.
XX 26-MAR-1996.
XX 23-DEC-1985; 85US-0812879.
XX 22-DEC-1986; 86US-0944117.
XX 23-DEC-1985; 85US-0812879.
XX (BEHW ) BEHRINGWERKE AG.
XX (CHIR ) CHIRON CORP.
XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
XX WPI; 1996-179247/18.
XX Human tissue plasminogen activator mutants for fibrin clot lysis
XX with improved plasminogenolytic specific activity, increased fibrin
XX dependence and decreased plasminogen inhibitor susceptibility
XX Claim 3; Page -: 8pp; English.
XX AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
XX that have improved plasminogen specific activity, increased fibrin
XX dependence and decreased plasminogen inhibitor susceptibility. The
XX mutants are useful in the lysis of fibrin clots and in the
XX prevention of blood clot formation by activating plasminogen. The
XX specific activity of tPA is enhanced by reducing the amount of
XX glycosylation of the protein by modifying the glycosylation sites at
XX positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
CC
```

CC specific activity of tPA is enhanced by reducing the amount of  
CC glycosylation of the protein by modifying the glycosylation sites at  
CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to  
CC plasminogen activator inhibition is reduced by modifying the cleavage  
CC site at position 274-278 and the fibrin dependence of tPA is increased  
CC by truncating the C-terminus.

XX  
SQ Sequence 527 AA;

Query Match 44.3%; Score 39; DB 17; Length 527;  
Best Local Similarity 44.4%; Pred. No. 1.7e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 115 nwessalaqkpycgrrpd 132

RESULT 27  
AAW54150  
ID AAW54150 standard; protein; 527 AA.  
XX  
AC AAW54150;  
XX  
DT 20-JUL-1998 (first entry)  
XX  
DE t-PA mutant (S127C, Y143H, C173R).  
XX  
KW Amino acid substitution; t-PA; vascular disorder; prevention;  
KW fibrin deposition; adhesion formation.  
XX  
OS Synthetic.

Key	Location/Qualifiers
FT Domain	1..44
FT Domain	/note= "Finger domain"
FT Domain	45..91
FT Domain	/note= "Growth factor domain"
FT Domain	92..173
FT Domain	/note= "Kringle-1 domain"
FT Domain	180..261
FT Domain	/note= "Kringle-2 domain"
FT Domain	264..527
FT misc_difference 127	/note= "Serine protease domain"
FT misc_difference 143	/note= "S changed from wt to C in mutant"
FT misc_difference 173	/note= "Y changed from wt to H in mutant"
FT misc_difference 173	/note= "C changed from wt to R in mutant"
XX	
PN	US5736135-A.
XX	
PD	07-APR-1998.
XX	
PF	13-FEB-1995; 95US-0389615.
XX	
PR	11-JUL-1991; 91US-0728456.
PR	26-JAN-1993; 93US-0008940.
PR	01-APR-1994; 94US-0221660.
PR	13-FEB-1995; 95US-0389615.
XX	
PA	(GETH ) GENENTECH INC.
XX	
PI	Goeddel DV, Leung DWH, Rice GC;
XX	
DR	WPI; 1998-239153/21.
XX	
PT	Mutant tissue plasminogen activator proteins - useful for treating
PT	vascular disorders, preventing tissue adhesion(s), etc.
XX	
PS	Claim 6; Page -; 24pp; English.

XX Mutant tissue plasminogen activator proteins (AAW54147-W54158) are  
CC created by single or multiple amino acid substitutions. Compositions  
CC containing the t-PA variant are used for treating vascular disorders, for  
CC preventing fibrin deposition or for preventing adhesion formation or  
CC reformation. Note: This sequence is not given in the specification but  
CC was created from the wild type by the indexer.

XX  
SQ Sequence 527 AA;

Query Match 44.3%; Score 39; DB 19; Length 527;  
Best Local Similarity 44.4%; Pred. No. 1.7e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 115 nwnssalaqkpycgrrpd 132

RESULT 28  
AAR13153  
ID AAR13153 standard; Protein; 558 AA.  
XX  
AC AAR13153;  
XX  
DT 29-AUG-1991 (first entry)  
XX  
DE T-PA with -ve charged finger and/or kringle domain (8).  
XX  
KW T-PA; tissue plasminogen activator; finger; kringle domain;  
KW thrombosis; lysis.

Key	Location/Qualifiers
FT Region	33..33
FT Region	/label= R1
FT Region	34..38
FT Region	/label= A2-6
FT Region	/note= "Tyr2-Cys6 of natural t-PA"
FT Region	39..61
FT Region	/label= R2
FT Region	62..154
FT Region	/label= A31-123
FT Region	/note= "Val31-Gln123 of natural t-PA"
FT Region	155..166
FT Region	/label= R3
FT Region	157..558
FT Region	/label= A136-527
FT Region	/note= "Leul36-Pro527 of natural t-PA"
XX	
PN	JP03061483-A.
XX	
PD	18-MAR-1991.
XX	
PF	26-JUN-1989; 89JP-0163599.
XX	
PR	27-JUN-1988; 88GB-0015246.
XX	
PA	(FUJI ) FUJISAWA PHARM KK.
XX	
DR	WPI; 1991-180545/25.
DR	N-PSDB; AAQ12071.
XX	
PT	New tissue plasminogen activator with negatively charged finger
PT	and/or kringle domain - is highly active and is useful as thrombosis-
PT	lysing agent
XX	
PS	Claim 1; Page 1; 29pp; Japanese.
XX	
CC	This sequence is an example of a generic formula. The activator
CC	is highly active in converting plasminogen to plasmin. It is
CC	useful as a thrombosis-lysing agent and is used in a pharmaceutical
CC	compsn. to treat diseases, such as cardiac infarction, cerebral



CC haemorrhage, heart stroke and pulmonary embolism.  
CC See AAQ12071-2, AAR13150-53 and AAR12366-67 and also J03061482-84.  
XX  
SQ Sequence 558 AA;

Query Match 44.3%; Score 39; DB 12; Length 558;  
Best Local Similarity 44.4%; Pred. NO. 1.8e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTOEE 16  
Db 146 nwnssalaqepysgeepd 163

RESULT 29  
AAR13151  
ID AAR13151 standard; Protein; 559 AA.  
XX  
AC AAR13151;  
XX  
DT 29-AUG-1991 (first entry)  
XX  
DE T-PA with -ve charged finger and/or kringle domain (4).  
XX  
DE T-PA; tissue plasminogen activator; finger; kringle domain;  
KW thrombosis; lysis.  
KW

XX Key Location/Qualifiers  
FH 33..33  
FT /label= R1  
FT 34..38  
FT /label= A2-6  
FT /note= "Tyr2-Cys6 of natural t-PA"  
FT 39..62  
FT /label= R2  
FT 63..155  
FT /label= A31-123  
FT /note= "Val31-Gln123 of natural t-PA"  
FT 156..167  
FT /label= R3  
FT 168..559  
FT /label= A136-527  
FT /note= "Leu136-Pro527 of natural t-PA"  
XX  
PN JP03061483-A.  
XX  
PD 18-MAR-1991.  
XX  
PF 26-JUN-1989; 89JP-0163599.  
XX  
PR 27-JUN-1988; 88GB-0015246.  
XX  
PA (FUJI ) FUJISAWA PHARM KK.  
XX  
XX WPI; 1991-180545/25.  
XX N-PSDB; AAQ12071.  
XX  
XX New tissue plasminogen activator with negatively charged finger  
XX and/or kringle domain - is highly active and is useful as thrombosis-  
XX lysing agent  
PS Claim 1; Page 1; 29pp; Japanese.  
XX  
XX This sequence is an example of a generic formula. The activator  
XX is highly active in converting plasminogen to plasmin. It is  
XX useful as a thrombosis-lysing agent and is used in a pharmaceutical  
XX compsn. to treat diseases, such as cardiac infarction, cerebral  
XX haemorrhage, heart stroke, and pulmonary embolism.  
XX See AAQ12071-2, AAR13150-53 and AAR12366-67 and also J03061482-84.  
XX  
SQ Sequence 559 AA;

Query Match 44.3%; Score 39; DB 12; Length 559;  
Best Local Similarity 44.4%; Pred. NO. 1.8e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTOEE 16  
Db 147 nwnssalaqepysgeepd 164

RESULT 30  
AAR12367  
ID AAR12367 standard; Protein; 561 AA.  
XX  
AC AAR12367;  
XX  
DT 29-AUG-1991 (first entry)  
XX  
DE T-PA with -ve charged finger and/or kringle domain (7).  
XX  
DE T-PA; tissue plasminogen activator; finger; kringle domain;  
KW thrombosis; lysis.  
KW

XX Key Location/Qualifiers  
FH 33..36  
FT /label= R1  
FT 37..41  
FT /label= A2-6  
FT /note= "Tyr2-Cys6 of natural t-PA"  
FT 42..64  
FT /label= R2  
FT 65..157  
FT /label= A31-123  
FT /note= "Val31-Gln123 of natural t-PA"  
FT 158..169  
FT /label= R3  
FT 170..561  
FT /label= A136-527  
FT /note= "Leu136-Pro527 of natural t-PA"  
XX  
PN JP03061483-A.  
XX  
PD 18-MAR-1991.  
XX  
PF 26-JUN-1989; 89JP-0163599.  
XX  
PR 27-JUN-1988; 88GB-0015246.  
XX  
PA (FUJI ) FUJISAWA PHARM KK.  
XX  
XX WPI; 1991-180545/25.  
XX N-PSDB; AAQ12071.  
XX  
XX New tissue plasminogen activator with negatively charged finger  
XX and/or kringle domain - is highly active and is useful as thrombosis-  
XX lysing agent  
PS Claim 1; Page 1; 29pp; Japanese.  
XX  
XX This sequence is an example of a generic formula. The activator  
XX is highly active in converting plasminogen to plasmin. It is  
XX useful as a thrombosis-lysing agent and is used in a pharmaceutical  
XX compsn. to treat diseases, such as cardiac infarction, cerebral  
XX haemorrhage, heart stroke, and pulmonary embolism.  
XX See AAQ12071-2, AAR13150-53 and AAR12366-67 and also J03061482-84.  
XX  
SQ Sequence 561 AA;

Query Match 44.3%; Score 39; DB 12; Length 561;  
Best Local Similarity 44.4%; Pred. NO. 1.8e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

```
.Qy 1 DWDTsALA--PYLGTQEE 16
      ::::|||||  ||| : :
Db 149 nwnssalagqepysqeed 166
```

RESULT 31  
AAR12343  
ID AAR12343 standard; Protein; 562 AA.

JP03061483-A.

18-MAR-1991.

26-JUN-1989; 89JP-0163599;

27-JUN-1988; 88GB-0015246.

(FUJI ) FUJISAWA PHARM KK.

WPI; 1991-180545/25.

N-PSDB; AAQ12071.

PT New tissue plasminogen activator with negatively charged finger  
PT and/or kringle domain - is highly active and is useful as thrombosis-  
PT lysing agent

Disclosure; Fiq 28; 29pp; Japanese.

This sequence is an example of a generic formula. The activator is highly active in converting plasminogen to plasmin. It is useful as a thrombosis-lysing agent and is used in a pharmaceutical compsn. to treat diseases, such as cardiac infarction, cerebral haemorrhage, heart stroke and pulmonary embolism.  
See AAQ12071-2, AAR13150-53 and AAR12366-67 and also J03061482-84.

Sequence 562 AA;

Query Match 44.3%; Score 39; DB 12; Length 562;  
Best Local Similarity 44.4%; Pred. NO. 1.8e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels

```
QY      1 DWDTSALA--PYLGTQEE 16
          :|:| | | | | | | :
Db     150 nwnssalaqepysqeepd 167
```

RESULT 32  
AAR20210  
ID AAR20210 standard; Protein; 734 AA.  
XX  
AC AAR20210;

Di:acyl:glycerol kinase which allows encoded cDNA to be isolated  
- in which DNA contains base sequence which encodes amino acid  
sequence of di:acyl:glycerol kinase activity

Claim 4; Page 1 and Fig 1; 13pp; Japanese.

Total mRNA was extracted from swine thymus cells. The cDNA was synthesised using oligo(dT) as primer and commercial BRL, fractionated by agarose gel filtration to settled size (e.g. at least 1 kb) and collected. A cDNA library was constructed in pBR322 by using dC-dG homopolymer tailing method. The DCK-encoding clones were isolated following hybridisation with a probe. The sequence of each fragment was determined to obtain the complete nucleotide sequence. See AAQ21834-35 for probes.

Sequence 734 AA;

Query Match 44.3%; Score 39; DB 13; Length 734;  
Best Local Similarity 63.6%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels

```
Qy 1 DWDTSALAPYL 11
    ||| ||| |||
Db 149 dwdvselrpil 159
```

RESULT 33  
AAG00538  
ID AAG00538 standard; Protein; 65 AA.  
XX  
AC AAG00538;

XX PF 21-FEB-2000; 2000EP-0200610.  
XX PR 26-FEB-1999; 99US-0122487.  
XX PA (GEST ) GENSET.  
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX DR WPI; 2000-500381/45.  
XX N-PSDB; AAC00544.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 13; SEQ ID 4619; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number  
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
XX were prepared from total human RNAs or polyA+ RNAs derived from 30  
XX different tissues. EST sequences usually correspond mainly to the 3'  
XX untranslated region (UTR) of the mRNA because they are often obtained  
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
XX those cases where longer cDNA sequences have been obtained, the full 5'  
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
XX ends and can therefore be used to obtain full length cDNAs and genomic  
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
XX chromosome mapping procedures. They are used to obtain upstream  
XX regulatory sequences and to design expression and secretion vectors.

XX Sequence 65 AA;

Query Match 43.2%; Score 38; DB 21; Length 65;  
Best Local Similarity 75.0%; Pred. No. 25;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WDTSALAP 9  
Db 18 wdtleslap 25

RESULT 34  
AAB08408  
ID AAB08408 standard; Protein; 87 AA.  
XX AC AAB08408;

XX 20-DEC-2000 (first entry)

XX Amino acid sequence of kringle 1 of human plasminogen activator.

XX Angiogenesis inhibitor; KED; kringle protein; angiogenesis; tPA;  
XX tissue plasminogen activator; tumour; atherosclerosis; arthritis;  
XX retinopathy; bronchial vascular congestion; inflammatory bowel disease;  
XX adult respiratory distress syndrome; Castlemans disease; psoriasis;  
XX hepatitis; aneurysm; renal disease; haemangioma.

XX Homo sapiens.

XX WO200049871-A1.

XX 31-AUG-2000.

XX 24-FEB-2000; 2000WO-US04798.

XX 24-FEB-1999; 99US-0121341.

XX 25-FEB-1999; 99US-0121633.

XX 18-NOV-1999; 99US-0166176.

XX (FORD-) FORD HEALTH SYSTEM HENRY.

XX Dou D, Chopp M, Wang L, Mikkelsen T;  
XX WPI; 2000-572016/53.  
XX Use of kringle protein and kringle derived from plasminogen and  
XX composition comprising kringle proteins for treating tumor and  
XX atherosclerosis, arthritis and retinopathy  
XX Disclosure; Fig 6; 163pp; English.

XX The specification describes a human polypeptide which is a potent  
XX angiogenesis inhibitor, and is designated KED. KED polypeptide, kringle  
XX proteins, or a kringle derived from human tissue plasminogen activator  
XX (tPA) protein are used to inhibit angiogenesis. Kringle protein, KED  
XX or tPA are useful for treating tumours, as well as atherosclerosis,  
XX arthritis, retinopathy and other similar diseases. KED is also useful  
XX for the treatment of diseases such as bronchial vascular congestion,  
XX inflammatory bowel disease, adult respiratory distress syndrome,  
XX Castlemans disease, psoriasis, hepatitis, aneurysm, renal disease  
XX and haemangioma. The present sequence represents kringle 1 of human  
XX tPA, which is used in the course of the invention.

XX Sequence 87 AA;

Query Match 43.2%; Score 38; DB 21; Length 87;  
Best Local Similarity 44.4%; Pred. No. 34;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGQTQEE 16  
Db 27 nwnssalaqkpysgrrpd 44

RESULT 35  
AAW13516  
ID AAW13516 standard; protein; 118 AA.

XX AAW13516;

XX 28-OCT-1997 (first entry)

XX Anti-melanoma antibody heavy chain clone D33.

XX Human; monoclonal antitumour antibody; peripheral blood lymphocyte;  
XX cancer; tumorigenesis; anticancer vaccine.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 76

XX /label= Leu, Thr

XX 30..34

XX /label= CDR1

XX 49..65

XX /label= CDR2

XX 98..109

XX /label= CDR3

XX WO9702479-A2.

XX 23-JAN-1997.

XX 28-JUN-1996; 96WO-IB01032.

XX 30-JUN-1995; 95US-0497647.

XX (UYVA ) UNIV YALE.

XX Cal X, Garen A;

XX WPI; 1997-109061/10.

XX Prodn. of human monoclonal anti-tumour antibodies - by screening a  
PT fusion phage library produced using peripheral blood lymphocytes  
PT from a cancer patient  
XX  
PS Claim 19; Page 54; 82pp; English.  
XX  
CC A process for isolating and synthesising human monoclonal anti-tumour  
CC antibodies has been produced. The process involves: (a) constructing at  
CC least one fusion phage library from the peripheral blood lymphocytes  
CC (PBLs) of a cancer patient; (b) screening for anti-tumour antibodies in  
CC the phage library in a binding assay with cultured tumour cells of the  
CC same type as the patient's tumour; (c) removing extraneous antibodies by  
CC absorption against normal human cells; (d) cloning the phage selected in  
CC step (b) and (c); (e) assaying the specificity of the cloned phage by  
CC incubating the phage with at least two types of cultured normal cells;  
CC and (f) further testing the specificity of cloned phage that do not bind  
CC to either cell line of cultured normal cells in further binding assays  
CC to cultured tumour cells derived from more than one other tumour that is  
CC not the patient's tumour. The present sequence represents a human heavy  
CC chain antibody, from an scFv antibody fusion phage library, produced by  
CC a method as described above. The antibodies produced can be used for  
CC diagnostic and therapeutic applications and for isolating tumour  
CC antigens for studying tumourigenesis or for use as anti-cancer vaccines.  
CC The human antibodies have low immunogenicity in humans compared to  
CC murine monoclonal antibodies (MAbs). Since the antibodies are isolated  
CC from fusion phage libraries, their affinity and specificity for a  
CC tumour cell line can be improved by genetic manipulations.  
XX  
SQ Sequence 118 AA;  
  
Query Match 43.2%; Score 38; DB 18; Length 118;  
Best Local Similarity 54.5%; Pred. No. 48;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 WDTSALAPYLG 12  
Db ||||| |:  
71 wdtsixtaymg 81  
  
RESULT 36  
AAW04249  
ID AAW04249 standard; Protein; 209 AA.  
XX  
AC AAW04249;  
XX  
DT 30-DEC-1996 (first entry)  
XX  
DE 4070A retrovirus amphotropic gp70 protein.  
XX  
KW Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;  
KW 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;  
KW polytropic MX27 provirus; targetted drug delivery; gene therapy;  
KW single chain antibody; envelope protein.  
XX  
OS 4070A retrovirus.  
XX  
PN WO9630504-A1.  
XX  
PD 03-OCT-1996.  
XX  
PF 22-MAR-1996; 96WO-US03908.  
XX  
PR 24-MAR-1995; 95US-0409648.  
XX  
PA (GENE-) GENETIC THERAPY INC.  
PA (UYSC-) UNIV SOUTHERN CALIFORNIA.  
XX  
PI Anderson W, Chiang YL, Januszeski M, Mackrell AJ;  
PI Zhao Y;  
XX  
DR WPI; 1996-455352/45.

DR N-PSDB; AAT33923.  
XX  
PT Cell-targetted retroviral vector particles - having envelope protein  
PT modified with targetting polypeptide  
XX  
PS Claim 1; Page 44-45; 73pp; English.  
XX  
CC Cell targetted retroviral vector particles can be used in gene  
CC therapy to deliver a heterologous gene to a target cell for  
CC expression of a heterologous polypeptide in that cell. The cell  
CC targetted retroviral vector particles comprise an envelope protein  
CC which is modified to contain a targetting polypeptide (a single chain  
CC antibody). The preferred modifications to this sequence comprise the  
CC replacement of amino acids 47-75, 47-93, 37-163.  
XX  
SQ Sequence 209 AA;  
  
Query Match 43.2%; Score 38; DB 17; Length 209;  
Best Local Similarity 50.0%; Pred. No. 91;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 DWDTSALAPYLG 12  
Db :||| | |:  
53 ewdpsdqepvvg 64  
  
RESULT 37  
AAW04250  
ID AAW04250 standard; Protein; 209 AA.  
XX  
AC AAW04250;  
XX  
DT 30-DEC-1996 (first entry)  
XX  
DE 10A1 murine leukaemia virus envelope protein.  
XX  
KW Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;  
KW 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;  
KW polytropic MX27 provirus; targetted drug delivery; gene therapy;  
KW single chain antibody; envelope protein.  
XX  
OS 10A1 murine leukaemia virus.  
XX  
PN WO9630504-A1.  
XX  
PD 03-OCT-1996.  
XX  
PF 22-MAR-1996; 96WO-US03908.  
XX  
PR 24-MAR-1995; 95US-0409648.  
XX  
PA (GENE-) GENETIC THERAPY INC.  
PA (UYSC-) UNIV SOUTHERN CALIFORNIA.  
XX  
PI Anderson W, Chiang YL, Januszeski M, Mackrell AJ;  
PI Zhao Y;  
XX  
DR WPI; 1996-455352/45.  
XX  
DR N-PSDB; AAT33924.  
XX  
PT Cell-targetted retroviral vector particles - having envelope protein  
PT modified with targetting polypeptide  
XX  
PS Claim 1; Page 45-47; 73pp; English.  
XX  
CC Cell targetted retroviral vector particles can be used in gene  
CC therapy to deliver a heterologous gene to a target cell for  
CC expression of a heterologous polypeptide in that cell. The cell  
CC targetted retroviral vector particles comprise an envelope protein  
CC which is modified to contain a targetting polypeptide (a single chain  
CC antibody). The preferred modifications to this sequence comprise the  
CC replacement of amino acids 47-75, 47-93, 47-163.



CC of (II) or for treatment of cancer pathologies (all claimed).

XX Sequence 209 AA;

Query Match 43.2%; Score 38; DB 17; Length 209;  
Best Local Similarity 50.0%; Pred. No. 91;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLG 12  
:||| | ||:|  
Db 53 ewdpsdqepvg 64

RESULT 38  
AAB19865  
ID AAB19865 standard; Protein; 288 AA.  
XX  
AC AAB19865;  
XX  
DT 19-MAR-2001 (first entry)  
XX  
DE Amphotropic MLV surface subunit RBD domain and PHQV motif.  
XX  
KW Amphotropic; MLV; surface subunit; receptor binding domain; RBD;  
KW PHQV motif; activating polypeptide; envelope glycoprotein;  
KW retrovirus; fusion defective; infection; cancer; therapy;  
KW virucide; cytostatic.  
XX  
OS Murine leukemia virus.  
XX  
PN WO2000071578-A2.  
XX  
PD 30-NOV-2000.  
XX  
PF 19-MAY-2000; 2000WO-EP04534.  
XX  
PR 20-MAY-1999; 99EP-0401220.  
XX  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Cosset F, Lavillette D;  
XX  
DR WPI; 2001-061346/07.  
DR N-PSDB; AAA89065.  
XX  
PT Use of an activating polypeptide comprising an amino acid sequence from  
PT a surface subunit of a viral envelope glycoprotein, useful for the  
PT rescue of fusion defective viruses or retroviral glycoproteins, and for  
PT treating cancer.  
XX  
PS Disclosure; Fig 6; 96pp; English.  
XX  
CC The present sequence is that an activating polypeptide (I),  
CC comprising the receptor binding domain (RBD) and the PHQV  
CC motif of the amphotropic murine leukemia virus surface subunit  
CC (SU), useful in claimed methods of the invention for the rescue of  
CC a fusion defective envelope glycoprotein (II). (I) can comprise a  
CC sequence corresponding to the SU of an envelope glycoprotein of a  
CC first virus or a first retrovirus, a sequence derived from SU, or  
CC a fragment of SU encompassing the RBD and the N-terminal part of SU.  
CC (II) belongs to a second virus or a second retrovirus. (I) may  
CC or different type from the first virus or retrovirus. (I) may  
CC recognise the same or different receptors on target cells as are  
CC recognised by (II), and (I) and/or (II) may be linked to a ligand  
CC that is specific to target cell receptors, e.g. receptors of tumour  
CC cells. (I) may be expressed by a host, such as a virus or cell, or  
CC by the target cells, transformed by nucleotide sequences coding for  
CC (I). (II) may be expressed by a host, such as a virus  
CC incorporating (II) in its envelope, or a cell transformed by  
CC nucleotide sequences coding for (II). (I) and (II) are used for the  
CC preparation of a drug for the treatment of cancer and infectious  
CC diseases, and as combined preparations for simultaneous, separate  
CC or sequential use for the rescue of the fusion defective property

XX Sequence 288 AA;

Query Match 43.2%; Score 38; DB 22; Length 288;  
Best Local Similarity 50.0%; Pred. NO. 1.3e+02;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLG 12  
:||| | ||:|  
Db 81 ewdpsdqepvg 92

RESULT 39  
AAP70183  
ID AAP70183 standard; Protein; 293 AA.  
XX  
AC AAP70183;  
XX  
DT 27-MAY-1991 (first entry)  
XX  
DE Sequence encoded by signal sequence of S.diastaticus DEX4  
DE glucoamylase gene.  
XX  
KW Yeast expression vector.  
XX  
OS Saccharomyces diastaticus.  
XX  
PN EP228254-A.  
XX  
PD 08-JUL-1987.  
XX  
PF 18-DEC-1986; 86EP-0309904.  
XX  
PR 18-DEC-1985; 85US-0810423.  
XX  
PA (BIOT-) BIOTECHNICA INT INC.  
XX  
PI Maine GT, Daves RS, Yocum RR;  
XX  
DR WPI; 1987-186757/27.  
DR N-PSDB; AAN70265.  
XX  
PT Vector for transforming yeast - having secretory signal-encoding  
PT sequence of glucoamylase gene from Saccharomyces diastaticus or  
PT S-cerevisiae  
XX  
PS Disclosure; Fig 3; 11pp; English.  
XX  
CC AAN70265 contains the preferred signal sequence of the claimed vector.  
CC The vector can be used to transform yeast cells to produce a desired  
CC protein, eg glucoamylase from A.niger which can be used in brewing,  
CC corn syrup prodn. and grain fermentation for distd. EtOH. prodn.  
XX  
CC Sequence 293 AA;  
XX  
Query Match 43.2%; Score 38; DB 8; Length 293;  
Best Local Similarity 46.7%; Pred. NO. 1.3e+02;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 WDTLSALAPYLGQTQEE 16  
||:|||| |:  
Db 76 wdssaleewlqrqkk 90

RESULT 40  
AAP80770  
ID AAP80770 standard; protein; 310 AA.  
XX  
AC AAP80770;  
XX

DT 31-OCT-1990 (first entry)  
XX  
DE Recombinant human tissue plasminogen activator A chain.  
XX  
KW human tissue plasminogen activator (t-PA) A chain;  
KW antithrombotic agents.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..35  
FT /label=signal peptide  
FT Protein 36..310  
FT /label=mature t-PA A-chain  
XX  
PN W08808878-A.  
XX  
PD 17-NOV-1988.  
XX  
PF 12-MAY-1988; 88WO-US01624.  
XX  
PR 15-MAY-1987; 87US-0050950.  
PR 12-MAY-1988; 88WO-US01624.  
XX  
PA (INTE-) INTEG GENETICS INC.  
XX  
PI Reddy VB;  
XX  
DR WPI; 1988-338227/47.  
DR N-PSDB; AAN80896.  
XX  
PT DNA coding for tissue plasminogen activator A chain -  
PT used as labelled A chain for antithrombotic drug screening  
XX  
PS Disclosure; ; p; English.  
XX  
CC Protein encoded by recombinant DNA which was obtained upon  
CC ligation of a 184bp Ddel-EcoRI fragment of p7BDT-10, a 730bp  
CC Sali-EcoRI fragment of p7BDT-10 encoding N-terminal portion of  
CC t-PA and a synthetic oligomer capable of being transcribed into  
CC an mRNA stop codon. The ligated product was cleaved with Sali  
CC to give the 952bp fragment shown. This fragment was cloned into  
CC the XhoI site of bovine papilloma virus DNA and used to transfect  
CC C127 mouse cells.  
CC The protein is labelled and used to screen for antithrombotic  
CC agents or to detect and monitor blood clots in vivo.  
XX  
SQ Sequence 310 AA;  
  
Query Match 43.2%; Score 38; DB 9; Length 310;  
Best Local Similarity 44.4%; Pred. No. 1.4e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;  
  
QY 1 DWDTSALA--PYLGTEEE 16  
:|::| | | | | | :  
Db 150 nwnssalagkypysgrpd 167  
  
RESULT 41  
AAP90174  
ID AAP90174 standard; peptide; 314 AA.  
XX  
AC AAP90174;  
XX  
DT 21-DEC-1989 (first entry)  
XX  
DE Tissue plasminogen activator mutant 2K1.  
XX  
KW Tissue plasminogen activator; mutant 2K1; fibrinolysis.  
XX  
OS Homo sapiens.  
XX

PN W08907146-A.  
XX  
PD 10-AUG-1989.  
XX  
PF 03-FEB-1989; 89WO-US00465.  
XX  
PR 05-FEB-1988; 88US-0152692.  
XX  
PA (INTE-) INTEGRATED GENETICS INC.  
XX  
PI Markland W, Livingston DJ;  
XX  
DR WPI; 1989-249015/34.  
DR N-PSDB; AAN90564.  
XX  
PT Rearranged tissue plasminogen activators - prepd. by altering the DNA  
PT sequence to introduce AvrII, NheI, SpeI or XbaI cleavage sites.  
XX  
PS Disclosure; Table 10; page 48; 77pp; English.  
XX  
CC Peptide has the kringle 2 domain replaced by another kringle 1 domain.  
CC It has tPA activity, with extended in vivo half life. Spacing between  
CC tPA domains is increased, to increase rate of fibrinolysis or the  
CC resistance to inhibition by endogenous tPA inhibitors present in human  
CC plasma. The peptide is used for thrombolysis in the treatment of  
CC myocardial infarction, pulmonary embolism, deep vein thrombosis and  
CC stroke. See also AAN90564, AAP90168-73, and AAP90175-80.  
XX  
SQ Sequence 314 AA;  
  
Query Match 43.2%; Score 38; DB 10; Length 314;  
Best Local Similarity 44.4%; Pred. No. 1.4e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;  
  
QY 1 DWDTSALA--PYLGTEEE 16  
:|::| | | | | | :  
Db 269 nwnssalagkypysgrpd 286  
  
RESULT 42  
AAW28537  
ID AAW28537 standard; Protein; 356 AA.  
XX  
AC AAW28537;  
XX  
DT 22-JAN-1998 (first entry)  
XX  
DE mt-PA9.  
XX  
KW Tissue plasminogen activator; chimera; kringle; mt-PA6-D;  
KW mt-PA6-E; mt-PA9.  
XX  
OS Homo sapiens.  
XX  
PN US5658788-A.  
XX  
PD 19-AUG-1997.  
XX  
PF 22-APR-1991; 91US-0689410.  
XX  
PR 22-APR-1991; 91US-0689410.  
PR 21-APR-1995; 95US-0427640.  
XX  
PA (ELIL ) LILLY & CO ELI.  
XX  
PI Berg DT, Grinnell BW;  
XX  
DR WPI; 1997-424241/39.  
DR N-PSDB; AAT87457.  
XX  
PT Recombinant DNA encoding tissue plasminogen activator variant -  
PT comprising signal and pro peptide(s) and chimeric kringle domain,

PT produced in di:glycosylated form, useful to treat thromboembolic  
PT diseases  
XX  
PS Disclosure; Col 43-46; 44pp; English.  
XX  
CC mt-PA6-D (AA787455) comprises an alternate glycosylation consensus  
CC sequence at the amino acids surrounding position 33 (equivalent  
CC to amino acid 205 of t-PA).  
CC mt-PA6-E (AA787456) comprises an alternate glycosylation consensus  
CC sequence at the amino acids surrounding position 48 (equivalent  
CC to amino acid 219 of t-PA).  
CC mt-PA9 (AA787457) is a human t-PA derivative that lacks the finger and  
CC growth factor domains but has a chimeric kringle domain having  
CC a part of kringle 1 that can be glycosylated, and a part of  
CC kringle 2 that cannot.  
XX  
SQ Sequence 356 AA;

Query Match 43.2%; Score 38; DB 18; Length 356;  
Best Local Similarity 44.4%; Pred. NO. 1.6e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTQEE 16  
Db 32 nwnssalaqkpysgripd 49

RESULT 43  
AAW14432  
ID AAW14432 standard; Protein; 356 AA.  
XX  
AC AAW14432;  
XX  
DT 14-MAY-1997 (first entry)  
XX  
DE Tissue plasminogen activator derivative, mt-PA9.  
DE  
DE Human tissue plasminogen activator; t-PA; derivative; mt-PA6; Finger;  
KW Growth Factor; Kringle 1; domain; Kringle 2; monoglycosylated; mt-PA6-E;  
KW Chimeric; mt-PA9; treatment; thromboembolic disorder; consensus sequence;  
KW alternate glycosylation; serine protease.  
XX  
XX Synthetic.  
XX  
XX US5595736-A.  
XX  
XX 21-JAN-1997.  
XX  
XX 22-APR-1991; 9LUS-0689410.  
XX  
XX 22-APR-1991; 9LUS-0689410.  
PR 27-MAR-1995; 9SUS-0411260.  
XX  
XX (ELIL ) LILLY & CO ELI.  
XX  
XX Berg DT, Grinnell BW;  
PI  
XX  
XX WPI; 1997-107532/10.  
DR N-PSDB; AAT62607.

XX Human tissue plasminogen activator derivs. - useful for treating  
PT thromboembolic disorders  
PT  
PT  
XX  
XX Claim 4; Column 43-46; 35pp; English.  
XX  
XX Human tissue plasminogen activator (t-PA) derivs. (especially derivs. of  
CC mt-PA6) have been designed which: (A) lack Finger, Growth Factor and  
CC Kringle 1 domains and comprises a Kringle 2 domain that is mono-  
CC glycosylated at an amino acid residue equiv. to amino acid 218 of native  
CC human t-PA; (B) lack Finger, Growth Factor and Kringle 1 domains and  
CC comprises a Kringle 2 domain that is monoglycosylated at amino acid 47  
CC (designated mt-PA6-E; AAW14431); and (C) lack Finger and Growth Factor

CC domains and has chimeric Kringle 1 and 2 domains (designated mt-PA9;  
CC AAW14432). The t-PA derivs. are useful for treating thromboembolic  
CC disorders. The alternate glycosylation sequence, as well as the  
CC glycosylation consensus sequence within the serine protease domain, is  
CC glycosylated upon the expression and secretion of these molecules from  
CC eukaryotic host cells. Thus, a homogeneous population of diglycosylated  
CC t-PA derivs. that lack the finger, growth factor and Kringle 1 domains,  
CC is produced. The present sequence is mt-PA9.  
XX  
SQ Sequence 356 AA;

Query Match 43.2%; Score 38; DB 18; Length 356;  
Best Local Similarity 44.4%; Pred. NO. 1.6e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTQEE 16  
Db 32 nwnssalaqkpysgripd 49

RESULT 44  
AAB19868  
ID AAB19868 standard; Protein; 364 AA.  
XX  
AC AAB19868;  
XX  
DT 19-MAR-2001 (first entry)  
XX  
DE Activating polypeptide A-RBD-epidermal growth factor fusion.  
DE  
DE Amphotropic; MLV; surface subunit; receptor binding domain; RBD;  
KW PHQV motif; activating polypeptide; envelope glycoprotein;  
KW retrovirus; fusion defective; infection; cancer; therapy;  
KW virucide; cytostatic; epidermal growth factor.  
XX  
XX Chimeric - Murine leukemia virus.  
OS Chimeric - Mammalia.  
XX  
XX WO2000071578-A2.  
XX  
XX 30-NOV-2000.  
XX  
XX 19-MAY-2000; 2000WO-EP04534.  
XX  
XX 20-MAY-1999; 99EP-0401220.  
XX  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Cosset F, Lavillette D;  
PI  
XX  
XX WPI; 2001-061346/07.  
DR N-PSDB; AAA89068.

Use of an activating polypeptide comprising an amino acid sequence from  
a surface subunit of a viral envelope glycoprotein, useful for the  
rescue of fusion defective viruses or retroviral glycoproteins, and for  
treating cancer -

Disclosure; Fig 9; 96pp; English.

The present sequence is that of a fusion protein between an  
activating polypeptide comprising the receptor binding domain  
(RBD) and the PHQV motif of the amphotropic murine leukemia virus  
surface subunit (SU), and epidermal growth factor (EGF). This is an  
example of activating polypeptides (I) used in the invention for the  
rescue of a fusion defective envelope glycoprotein (II). (I) may  
comprise a sequence corresponding to the SU of an envelope  
glycoprotein of a first virus or a first retrovirus, a sequence  
derived from SU, or a fragment of SU encompassing the RBD and the  
N-terminal part of SU. (II) belongs to a second virus or a second  
retrovirus of the same or different type from the first virus or  
retrovirus. (I) may recognise the same or different receptors on

CC target cells as are recognised by (II), and (I) and/or (II) may be  
CC linked to a ligand that is specific to target cell receptors, e.g.  
CC receptors of tumour cells such as EGF. (I) may be expressed by a  
CC host, such as a virus or cell, or by the target cells, transformed  
CC by nucleotide sequences coding for (I). (II) may be expressed by a  
CC host, such as a virus incorporating (II) in its envelope, or a cell  
CC transformed by nucleotide sequences coding for (II). (I) and (II)  
CC are used for the preparation of a drug for the treatment of cancer  
CC and infectious diseases, and as combined preparations for  
CC simultaneous, separate or sequential use for the rescue of the  
CC fusion defective property of (II) or for treatment of cancer  
CC pathologies (all claimed).

XX SQ Sequence 364 AA;

Query Match 43.2%; Score 38; DB 22; Length 364;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGL 12  
Db :||| :|||  
157 ewdpsdqepyg 168

RESULT 45  
AAR67652  
ID AAR67652 standard; Protein; 409 AA.  
XX AC AAR67652;  
XX 16-AUG-1995 (first entry)  
DT Glucomyrase.  
XX Aqualysin I; glucomyrase; extracellular production.  
KW Saccharomyces cerevisiae.  
XX  
OS  
XX  
XX Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= sig\_peptide  
XX  
PN JP06303984-A.  
XX  
PD 01-NOV-1994.  
XX  
PF 19-APR-1993; 93JP-0115378.  
XX  
PR 19-APR-1993; 93JP-0115378.  
XX  
PA (NISB ) JAPAN TOBACCO INC.  
XX  
XX WPI; 1995-018277/03.  
DR N-PSDB; AAQ75858.  
XX  
PT Recombinant DNA having sequence deleted from the aqua-lysin I  
PT precursor gene - for expression of mature aqua-lysin I in yeast  
XX  
PS Claim 1; Pages 7-8; 16pp; Japanese.  
XX  
CC AAQ75858 is the yeast glucomyrase (AAR67652) DNA, into which the  
CC aqualysin I (AAR67653) DNA (AAQ75859) minus its C-terminal pro  
CC sequence can be inserted. Resulting in the extracellular  
CC production of aqualysin I in yeast.  
XX

SQ Sequence 409 AA;

Query Match 43.2%; Score 38; DB 16; Length 409;  
Best Local Similarity 46.7%; Pred. No. 1.9e+02;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTLSALAPYLGTQEE 16  
Db :||| :|||  
333 wdssaleewlqrqkk 347  
  
Search completed: June 28, 2001, 11:54:15  
Job time: 76 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:11 ; Search time 37.02 Seconds  
(without alignments)  
8.706 Million cell updates/sec

Title: US-09-439-313-562  
Perfect score: 88  
Sequence: 1 DWDTALAPVLGTQEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCRVUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	52.3	225	2	US-08-637-759B-127
2	46	52.3	225	3	US-08-871-355A-127
3	41	46.6	233	2	US-08-928-692-31
4	39	44.3	270	4	US-09-362-473-10
5	39	44.3	429	4	US-09-362-473-12
6	38	43.2	43	4	US-08-930-503A-17
7	38	43.2	83	2	US-08-811-949-2
8	38	43.2	118	4	US-08-983-607-24
9	38	43.2	356	1	US-08-427-640-8
10	38	43.2	437	2	US-08-811-949-49
11	38	43.2	437	2	US-08-811-949-51
12	38	43.2	437	2	US-08-811-949-55
13	38	43.2	437	2	US-08-811-949-57
14	38	43.2	472	2	US-08-811-949-63
15	38	43.2	527	1	US-07-609-510B-16
16	38	43.2	527	2	US-08-811-949-39
17	38	43.2	527	5	PCT-US91-01025A-2
18	38	43.2	527	6	5185259-8
19	38	43.2	527	6	5520913-1
20	38	43.2	546	6	5200340-6
21	38	43.2	562	2	US-08-811-949-43
22	38	43.2	562	2	US-08-560-098A-50
23	38	43.2	562	2	US-08-883-795A-38
24	38	43.2	562	6	5185259-3
25	38	43.2	562	6	5200340-2
26	38	43.2	562	6	5344773-2
27	38	43.2	609	3	US-08-941-445A-7

28	38	43.2	806	1	US-08-270-076A-11	Sequence 11, Appl
29	38	43.2	912	5	PCT-US95-03747-2	Sequence 2, Appli
30	37.5	42.6	260	2	US-08-864-799-3	Sequence 3, Appli
31	37	42.0	121	4	US-08-983-607-20	Sequence 20, Appl
32	37	42.0	200	1	US-08-233-146-1	Sequence 1, Appli
33	37	42.0	200	1	US-08-463-470-1	Sequence 1, Appli
34	37	42.0	442	3	US-08-993-359-26	Sequence 26, Appl
35	36	40.9	588	2	US-08-620-605D-2	Sequence 2, Appli
36	36	40.9	588	2	US-09-005-232A-2	Sequence 2, Appli
37	36	40.9	595	1	US-08-225-989-2	Sequence 2, Appli
38	36	40.9	595	1	US-08-570-923-2	Sequence 2, Appli
39	36	40.9	595	1	US-08-580-014-2	Sequence 2, Appli
40	36	40.9	595	2	US-08-232-087A-2	Sequence 2, Appli
41	36	40.9	595	4	US-09-079-785-2	Sequence 2, Appli
42	35	39.8	149	2	US-08-815-175-3	Sequence 3, Appli
43	35	39.8	205	2	US-08-852-809-5	Sequence 5, Appli
44	35	39.8	316	1	US-08-414-926A-22	Sequence 22, Appl
45	35	39.8	316	2	US-08-926-922-22	Sequence 22, Appl
46	35	39.8	316	3	US-09-253-682-22	Sequence 22, Appl
47	35	39.8	347	3	US-09-059-369-2	Sequence 2, Appli
48	35	39.8	456	2	US-08-860-150-2	Sequence 2, Appli
49	35	39.8	456	3	US-09-338-132-2	Sequence 2, Appli
50	34	38.6	16	1	US-08-446-692-44	Sequence 44, Appl
51	34	38.6	16	2	US-08-488-351A-44	Sequence 44, Appl
52	34	38.6	16	5	PCT-US95-13841-16	Sequence 16, Appl
53	34	38.6	30	1	US-08-446-692-23	Sequence 23, Appl
54	34	38.6	30	2	US-08-488-351A-23	Sequence 23, Appl
55	34	38.6	38	1	US-08-446-692-59	Sequence 59, Appl
56	34	38.6	38	1	US-08-446-692-65	Sequence 65, Appl
57	34	38.6	38	2	US-08-488-351A-59	Sequence 59, Appl
58	34	38.6	38	2	US-08-488-351A-65	Sequence 65, Appl
59	34	38.6	48	1	US-08-446-692-37	Sequence 37, Appl
60	34	38.6	48	2	US-08-488-351A-37	Sequence 37, Appl
61	34	38.6	189	4	US-08-816-346-8	Sequence 8, Appli
62	34	38.6	189	4	US-09-335-411-8	Sequence 8, Appli
63	34	38.6	234	4	US-09-354-129-14	Sequence 14, Appl
64	34	38.6	236	3	US-08-886-269-5	Sequence 5, Appli
65	34	38.6	236	3	US-09-167-647-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-08-637-759B-127  
; Sequence 127, Application US/08637759B  
; Patent No. 5876931  
; GENERAL INFORMATION:  
; APPLICANT: David William Holden  
; TITLE OF INVENTION: Identification of Genes  
; NUMBER OF SEQUENCES: 501  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/637,759B  
; FILING DATE: 03-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB95/02875  
; FILING DATE: 11-DEC-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPMS 101  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 127:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-637-759B-127

Query Match 52.3%; Score 46; DB 2; Length 225;  
Best Local Similarity 70.0%; Pred. No. 1;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DWDTSALAPY 10  
Db 102 EWDSKALAPY 111

RESULT 2  
US-08-871-355A-127  
Sequence 127, Application US/08871355A  
Patent No. 6015669  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS: 501  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,355A  
FILING DATE: 09-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPMS 101 CON  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 127:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-871-355A-127

Query Match 52.3%; Score 46; DB 3; Length 225;  
Best Local Similarity 70.0%; Pred. No. 1;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DWDTSALAPY 10  
Db 102 EWDSKALAPY 111

RESULT 3  
US-08-928-692-31  
Sequence 31, Application US/08928692  
Patent No. 5958727  
GENERAL INFORMATION:  
APPLICANT: Brody, Howard  
APPLICANT: Yaver, Deborah S.  
APPLICANT: Lamsa, Michael  
APPLICANT: Hansen, Kim  
TITLE OF INVENTION: Methods for Modifying the Production of  
TITLE OF INVENTION: a polypeptide  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 59587270 No. 5958727disk of No. 5958727th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,692  
FILING DATE: 12-SEPT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4944.200-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 233 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: NO. 5958727e  
US-08-928-692-31

Query Match 46.6%; Score 41; DB 2; Length 233;  
Best Local Similarity 53.8%; Pred. No. 8.4;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTLSALAPYLGTO 14  
Db 35 WDFGALEPYISGQ 47

RESULT 4  
US-09-362-473-10  
Sequence 10, Application US/09362473  
Patent No. 6218169  
GENERAL INFORMATION:  
APPLICANT: Cahoon, Edgar B. E.  
APPLICANT: Cahoon, Rebecca E.  
APPLICANT: Falco, S. Carl  
APPLICANT: Morgante, Michele  
APPLICANT: Rafalski, J. Antoni

APPLICANT: Hitz, William D.  
APPLICANT: Kinney, Anthony J.  
TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes  
FILE REFERENCE: BB-1197  
CURRENT APPLICATION NUMBER: US/09/362,473  
CURRENT FILING DATE: 1999-07-28  
EARLIER APPLICATION NUMBER: 60/094,783  
EARLIER FILING DATE: JULY 31, 1998  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 10  
LENGTH: 270  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-362-473-10

Query Match 44.3%; Score 39; DB 4; Length 270;  
Best Local Similarity 54.5%; Pred. NO. 23;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSALAPYL 12  
|:| | |:  
Db 97 WETPLGPFLG 107

RESULT 5  
US-09-362-473-12  
Sequence 12, Application US/09362473  
Patent No. 6218169  
GENERAL INFORMATION:  
APPLICANT: Cahoon, Edgar B.  
APPLICANT: Cahoon, Rebecca E.  
APPLICANT: Falco, S. Carl  
APPLICANT: Morgante, Michele  
APPLICANT: Rafalski, J. Antoni  
APPLICANT: Hitz, William D.  
APPLICANT: Kinney, Anthony J.  
TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes  
FILE REFERENCE: BB-1197  
CURRENT APPLICATION NUMBER: US/09/362,473  
CURRENT FILING DATE: 1999-07-28  
EARLIER APPLICATION NUMBER: 60/094,783  
EARLIER FILING DATE: JULY 31, 1998  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 12  
LENGTH: 429  
TYPE: PRT  
ORGANISM: Oryza sativa  
US-09-362-473-12

Query Match 44.3%; Score 39; DB 4; Length 429;  
Best Local Similarity 54.5%; Pred. NO. 39;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSALAPYL 12  
|:| | |:  
Db 255 WETPLGPFLG 265

RESULT 6  
US-08-930-503A-17  
Sequence 17, Application US/08930503A  
Patent No. 6132731  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: MURINE LEUKEMIA VIRUS VECTORS (AS AMENDED)  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.  
STREET: 2033 K Street, N.W. - Suite 800

CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/930,503A  
FILING DATE: October 8, 1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9506782.3  
FILING DATE: 01-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/GB96/00776  
FILING DATE: 01-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-930-503A-17

Query Match 43.2%; Score 38; DB 4; Length 43;  
Best Local Similarity 50.0%; Pred. NO. 4;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 12  
|:| | |:  
Db 4 EWDPSDQEPYVG 15

RESULT 7  
US-08-811-949-2  
Sequence 2, Application US/08811949  
Patent No. 5840533  
GENERAL INFORMATION:  
APPLICANT: NIWA, MINEO  
APPLICANT: SAITO, YOSHIMASA  
APPLICANT: SASAKI, HITOSHI  
APPLICANT: HAYASHI, MASAKO  
APPLICANT: NOTANI, JOUJI  
APPLICANT: KOBAYASHI, MASAKAZU  
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:



APPLICATION NUMBER: US/08/811,949  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 83 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-811-949-2

Query Match 43.2%; Score 38; DB 2; Length 83;  
Best Local Similarity 44.4%; Pred. No. 8.6;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 24 NWNSSALAQKPYSGRRPD 41

RESULT 8  
US-08-983-607-24  
Sequence 24, Application US/08983607  
Patent No. 6140470  
GENERAL INFORMATION:  
APPLICANT: Alan Garen  
APPLICANT: Xiaohong Cai  
TITLE OF INVENTION: Human Anti-Tumor Monoclonal Anti-  
bodies  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Department of Molecular Biophysics  
and Biochemistry, Yale University  
STREET: 266 Whitney Avenue  
CITY: New Haven  
STATE: Connecticut  
COUNTRY: United States of America  
ZIP: 06520-8114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" 1.44 Mb diskette  
COMPUTER: IBM PC  
OPERATING SYSTEM: MS DOS  
SOFTWARE: Word Processing  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/983,607  
FILING DATE: April 27, 1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/IB96/01032  
FILING DATE: June 28, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary M. Krinsky  
REGISTRATION NUMBER: 32423  
REFERENCE/DOCKET NUMBER: OCR-679  
TELEPHONE: 203-773-9544  
TELEFAX: 203-773-1183  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 118 residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE:  
DESCRIPTION: polypeptide  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens (melanoma patient immu-  
nized with autologous tumor cells)  
INDIVIDUAL ISOLATE: peripheral blood lymphocytes  
IMMEDIATE SOURCE:  
LIBRARY: DM414 scfv antibodies obtained from  
LIBRARY: FUSE5 fusion phage construct  
CLONE: D33  
FEATURE:  
NAME/KEY: heavy chain  
OTHER INFORMATION: Xaa in position 81 is Leu or  
OTHER INFORMATION: Thr  
US-08-983-607-24

Query Match 43.2%; Score 38; DB 4; Length 118;  
Best Local Similarity 54.5%; Pred. No. 13;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTALAPYLG 12  
Db 71 WDTSIXTAYMG 81

RESULT 9  
US-08-427-640-8  
Sequence 8, Application US/08427640  
Patent No. 5658788  
GENERAL INFORMATION:  
APPLICANT: Berg et al.  
TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disor-  
ders  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/427,640  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/689,410  
FILING DATE: 22 APRIL 1991  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 356 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-427-640-8

Query Match 43.2%; Score 38; DB 1; Length 356;  
Best Local Similarity 44.4%; Pred. No. 47;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 32 NWNSSALAQKPYSGRRPD 49

RESULT 10  
US-08-811-949-49

; Sequence 49, Application US/08811949  
; Patent No. 5840533  
; GENERAL INFORMATION:  
; APPLICANT: NIWA, MINEO  
; APPLICANT: SAITO, YOSHIMASA  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: HAYASHI, MASAKO  
; APPLICANT: NOTANI, JOUJI  
; APPLICANT: KOBAYASHI, MASAKAZU  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,949  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 18-966-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 437 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-949-49

Query Match 43.2%; Score 38; DB 2; Length 437;  
Best Local Similarity 44.4%; Pred. NO. 60;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::||| ||| : :  
Db 25 NWNSSALAKPKYSGRRPD 42  
RESULT 11  
US-08-811-949-51  
; Sequence 51, Application US/08811949  
; Patent No. 5840533  
; GENERAL INFORMATION:  
; APPLICANT: NIWA, MINEO  
; APPLICANT: SAITO, YOSHIMASA  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: HAYASHI, MASAKO  
; APPLICANT: NOTANI, JOUJI  
; APPLICANT: KOBAYASHI, MASAKAZU  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA

; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,949  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 18-966-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 437 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-949-51

Query Match 43.2%; Score 38; DB 2; Length 437;  
Best Local Similarity 44.4%; Pred. NO. 60;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::||| ||| : :  
Db 25 NWNSSALAKPKYSGRRPD 42

RESULT 12  
US-08-811-949-55  
; Sequence 55, Application US/08811949  
; Patent No. 5840533  
; GENERAL INFORMATION:  
; APPLICANT: NIWA, MINEO  
; APPLICANT: SAITO, YOSHIMASA  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: HAYASHI, MASAKO  
; APPLICANT: NOTANI, JOUJI  
; APPLICANT: KOBAYASHI, MASAKAZU  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,949  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 18-966-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-949-55

Query Match 43.2%; Score 38; DB 2; Length 437;  
Best Local Similarity 44.4%; Pred. NO. 60;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 25 WNWSSALAKPYSGRRPD 42

RESULT 13  
US-08-811-949-57  
Sequence 57, Application US/08811949  
Patent No. 5840533  
GENERAL INFORMATION:  
APPLICANT: NIWA, MINEO  
APPLICANT: SAITO, YOSHIMASA  
APPLICANT: SASAKI, HITOSHI  
APPLICANT: HAYASHI, MASAKO  
APPLICANT: NOTANI, JOUJI  
APPLICANT: KOBAYASHI, MASAKAZU  
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,949  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-949-57

Query Match 43.2%; Score 38; DB 2; Length 437;  
Best Local Similarity 44.4%; Pred. NO. 60;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 25 WNWSSALAKPYSGRRPD 42

RESULT 14

US-08-811-949-63  
Sequence 63, Application US/08811949  
Patent No. 5840533  
GENERAL INFORMATION:  
APPLICANT: NIWA, MINEO  
APPLICANT: SAITO, YOSHIMASA  
APPLICANT: SASAKI, HITOSHI  
APPLICANT: HAYASHI, MASAKO  
APPLICANT: NOTANI, JOUJI  
APPLICANT: KOBAYASHI, MASAKAZU  
TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,949  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 63:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 472 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-949-63

Query Match 43.2%; Score 38; DB 2; Length 472;  
Best Local Similarity 44.4%; Pred. NO. 65;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 60 WNWSSALAKPYSGRRPD 77

RESULT 15  
US-07-609-510B-16  
Sequence 16, Application US/07609510B  
Patent No. 5326700  
GENERAL INFORMATION:  
APPLICANT: Berg et al.  
TITLE OF INVENTION: Method for Altering Post-Translational Processing of Tissue  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN.  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh

;; SOFTWARE: Microsoft Word  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/609,510B  
;; FILING DATE: 19901106  
;; CLASSIFICATION: 435  
;; INFORMATION FOR SEQ ID NO: 16:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 527 amino acids  
;; TYPE: AMINO ACID  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-07-609-510B-16

Query Match 43.2%; Score 38; DB 1; Length 527;  
Best Local Similarity 44.4%; Pred. No. 74;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::| | | | :  
Db 115 NWNSSALAKPKYSGRRPD 132

RESULT 16  
US-08-811-949-39  
;; Sequence 39, Application US/08811949  
;; Patent No. 5840533  
;; GENERAL INFORMATION:  
;; APPLICANT: NIWA, MINEO  
;; APPLICANT: SAITO, YOSHIMASA  
;; APPLICANT: SASAKI, HITOSHI  
;; APPLICANT: HAYASHI, MASAKO  
;; APPLICANT: NOTANI, JOUJI  
;; APPLICANT: KOBAYASHI, MASAKAZU  
;; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
;; NUMBER OF SEQUENCES: 67  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
;; ADDRESSEE: P.C.  
;; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
;; CITY: ARLINGTON  
;; STATE: VA  
;; COUNTRY: USA  
;; ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,949  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 527 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-811-949-39

Query Match 43.2%; Score 38; DB 2; Length 527;  
Best Local Similarity 44.4%; Pred. No. 74;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::| | | | :  
Db 115 NWNSSALAKPKYSGRRPD 132

RESULT 17  
PCT-US91-01025A-2  
;; Sequence 2, Application PC/TUS9101025A  
;; GENERAL INFORMATION:  
;; APPLICANT: Genentech, Inc.  
;; TITLE OF INVENTION: Tissue Plasminogen Activator Having Fibrin  
;; TITLE OF INVENTION: Specific Properties  
;; NUMBER OF SEQUENCES: 2  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genentech, Inc.  
;; STREET: 460 Point San Bruno Blvd  
;; CITY: South San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01025A  
FILING DATE: 19910214  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/486,657  
FILING DATE: 1 March 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 454P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 527 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
PCT-US91-01025A-2

Query Match 43.2%; Score 38; DB 5; Length 527;  
Best Local Similarity 44.4%; Pred. No. 74;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::| | | | :  
Db 115 NWNSSALAKPKYSGRRPD 132

RESULT 18  
5185259-8  
;; Patent No. 5185259  
;; APPLICANT: GOEDEL, DAVID V.; KOHR, WILLIAM J.; PENNICA, DIANE;  
;; VEHAR, GORDON A.  
;; TITLE OF INVENTION: TRUNCATED HUMAN TISSUE PLASMINOGEN  
;; ACTIVATOR  
;; NUMBER OF SEQUENCES: 15  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/489,855  
;; FILING DATE: 02-MAR-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 12,694  
;; FILING DATE: 09-FEB-1987  
;; APPLICATION NUMBER: 483,052



```

; FILING DATE: 07-APR-1983
; APPLICATION NUMBER: 398,003
; FILING DATE: 14-JUL-1982
; APPLICATION NUMBER: 374,860
; FILING DATE: 05-MAY-1982
; SEQ ID NO:8:
; LENGTH: 527
5185259-8

Query Match 43.2%; Score 38; DB 6; Length 527;
Best Local Similarity 44.4%; Pred. No. 74;
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
:|:||||| | | :
Db 115 NWNSSALAKPYSGRRPD 132

RESULT 19
5520913-1
; Patent No. 5520913
; APPLICANT: ANDERSON, STEPHEN; BENNETT, WILLIAM F.; BOTSTEIN,
; DAVID; HIGGINS, DEBORAH L.; PAONI, NICHOLAS F.; ZOLLER, MARK J.
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR HAVING
; ZYMOGENIC PROPERTIES
; NUMBER OF SEQUENCES: 35
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/88,451
; FILING DATE: 06-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 770,510
; FILING DATE: 03-OCT-1991
; APPLICATION NUMBER: 384,608
; FILING DATE: 24-JUL-1989
; APPLICATION NUMBER: 240,856
; FILING DATE: 02-SEP-1988
; SEQ ID NO:1:
; LENGTH: 527
5520913-1

Query Match 43.2%; Score 38; DB 6; Length 527;
Best Local Similarity 44.4%; Pred. No. 74;
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
:|:||||| | | :
Db 115 NWNSSALAKPYSGRRPD 132

RESULT 20
5200340-6
; Patent No. 5200340
; APPLICANT: FOSTER, DONALD C.; MULVIHILL, EILEEN R.; O'HARA,
; PATRICK J.; PINGEL, KURT; YOSHITAKE, SHINJI
; TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN
; ACTIVATORS
; NUMBER OF SEQUENCES: 34
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/53,412
; FILING DATE: 22-MAY-1987
; SEQ ID NO:6:
; LENGTH: 546
5200340-6

Query Match 43.2%; Score 38; DB 6; Length 546;
Best Local Similarity 44.4%; Pred. No. 77;
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
:|:||||| | | :

```

```

Db 150 NWNSSALAKPYSGRRPD 167

RESULT 21
US-08-811-949-43
; Sequence 43, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; STREET: ARLINGTON
; CITY: VA
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-43

Query Match 43.2%; Score 38; DB 2; Length 562;
Best Local Similarity 44.4%; Pred. No. 80;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
:|:||||| | | :
Db 150 NWNSSALAKPYSGRRPD 167

RESULT 22
US-08-560-098A-50
; Sequence 50, Application US/08560098A
; Patent No. 5976841
; GENERAL INFORMATION:
; APPLICANT: WENNDT, Stephan
; APPLICANT: HEINZEL-WIELAND, Regina
; APPLICANT: STEFFENS, Gerd Josef
; TITLE OF INVENTION: Proteins having Fibrinolytic and
; TITLE OF INVENTION: Coagulation-inhibiting Properties
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington

```

; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/560,098A  
; FILING DATE: 17-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: P 44 40 892.7  
; FILING DATE: 17-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: EVANS, Joseph D.  
; REGISTRATION NUMBER: 26,269  
; REFERENCE/DOCKET NUMBER: 148/42448  
; TELEPHONE: (202) 628-8800  
; TELEFAX: (202) 628-8844  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-560-098A-50

Query Match 43.2%; Score 38; DB 2; Length 562;  
Best Local Similarity 44.4%; Pred. No. 80;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::||| | | :  
Db 150 NWNSSALAKPYSGRRPD 167

RESULT 23  
US-08-883-795A-38  
; Sequence 38, Application US/08883795A  
; Patent No. 5985607  
; GENERAL INFORMATION:  
; APPLICANT: Delcuve, Genevieve  
; APPLICANT: Awang, Gregor  
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERESKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/883,795A  
; FILING DATE: 27-JUN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 7841-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 364-7311  
; TELEFAX: (416) 361-1398

; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 562 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein (tpa)  
; US-08-883-795A-38

Query Match 43.2%; Score 38; DB 2; Length 562;  
Best Local Similarity 44.4%; Pred. No. 80;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::||| | | :  
Db 150 NWNSSALAKPYSGRRPD 167

RESULT 24  
5185259-3  
; Patent No. 5185259  
; APPLICANT: GOEDEL, DAVID V.; KOHR, WILLIAM J.; PENNICA, DIANE;  
; VEHR, GORDON A.  
; TITLE OF INVENTION: TRUNCATED HUMAN TISSUE PLASMINOGEN  
; ACTIVATOR  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/489,855  
; FILING DATE: 02-MAR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 12,694  
; FILING DATE: 09-FEB-1987  
; APPLICATION NUMBER: 483,052  
; FILING DATE: 07-APR-1983  
; APPLICATION NUMBER: 398,003  
; FILING DATE: 14-JUL-1982  
; APPLICATION NUMBER: 374,860  
; FILING DATE: 05-MAY-1982  
; SEQ ID NO: 3:  
; LENGTH: 562  
5185259-3

Query Match 43.2%; Score 38; DB 6; Length 562;  
Best Local Similarity 44.4%; Pred. No. 80;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
:|::||| | | :  
Db 150 NWNSSALAKPYSGRRPD 167

RESULT 25  
5200340-2  
; Patent No. 5200340  
; APPLICANT: FOSTER, DONALD C.; MULVIHILL, EILEEN R.; O'HARA,  
; PATRICK J.; PINGEL, KURT; YOSHITAKE, SHINJI  
; TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN  
; ACTIVATORS  
; NUMBER OF SEQUENCES: 34  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/53,412  
; FILING DATE: 22-MAY-1987  
; SEQ ID NO: 2:  
; LENGTH: 562  
5200340-2

Query Match 43.2%; Score 38; DB 6; Length 562;  
Best Local Similarity 44.4%; Pred. No. 80;  
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
5344773-2  
Db 150 WNSSALAQKPYSGRRPD 167

RESULT 26  
5344773-2  
; Patent No. 5344773  
; APPLICANT: WEI, CHA-MER,HSIUNG, NANCY;REDDY, VERMURI B.;  
; LEMONTT, JEFFREY F.;DACKOWSKI, WILLIAM;DOUGLAS, RICHARD;  
; COLE, EDWARD S.;PURCELL JR., RICHARD D.;LAU, DAVID TAI-YUI  
; TITLE OF INVENTION: HUMAN UTERINE TISSUE PLASMINOGEN  
; ACTIVATOR PRODUCED BY RECOMBIANT DNA  
; NUMBER OF SEQUENCES: 6  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/782,686  
; FILING DATE: 01-OCT-1985  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 656,770  
; FILING DATE: 01-OCT-1984  
; SEQ ID NO:2:  
; LENGTH: 562  
5344773-2

Query Match 43.2%; Score 38; DB 6; Length 562;  
Best Local Similarity 44.4%; Pred. No. 80;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
Db 150 WNSSALAQKPYSGRRPD 167

RESULT 27  
US-08-941-445A-7  
; Sequence 7, Application US/08941445A  
; Patent No. 6107060  
; GENERAL INFORMATION:  
; APPLICANT: Keeling, Peter  
; APPLICANT: Guan, Hanping  
; TITLE OF INVENTION: Starch Encapsulation  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/941,445A  
; FILING DATE: 30-SEP-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,855  
; FILING DATE: 30-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winner, Ellen P  
; REGISTRATION NUMBER: 28,547  
; REFERENCE/DOCKET NUMBER: 89-97  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 609 amino acids  
; TYPE: amino acid

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-941-445A-7

Query Match 43.2%; Score 38; DB 3; Length 609;  
Best Local Similarity 63.6%; Pred. No. 88;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11  
Db 234 DWHTGPLASYL 244

RESULT 28  
US-08-270-076A-11  
; Sequence 11, Application US/08270076A  
; Patent No. 5667986  
; GENERAL INFORMATION:  
; APPLICANT: Sleep, Darrell  
; APPLICANT: Goodey, Andrew R  
; APPLICANT: Vakeria, Diana  
; TITLE OF INVENTION: Yeast Promoter  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The BOC Group, Inc.  
; STREET: 100 Mountain Avenue, Murray Hill  
; CITY: New Providence  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07974  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/270,076A  
; FILING DATE: 01-JUL-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8923521.2  
; FILING DATE: 18-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/597,687  
; FILING DATE: 16-OCT-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/925,286  
; FILING DATE: 04-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Swope, R. Hain  
; REGISTRATION NUMBER: 24864  
; REFERENCE/DOCKET NUMBER: 92H834-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908/771-6292  
; TELEFAX: 908/771-6159  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 806 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-270-076A-11

Query Match 43.2%; Score 38; DB 1; Length 806;  
Best Local Similarity 46.7%; Pred. No. 1-2e+02;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16  
Db 333 WDSSALEWLRQKK 347





```

; LENGTH: 121 residues
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; DESCRIPTION: polypeptide
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens (melanoma patient immu-
; INDIVIDUAL ISOLATE: peripheral blood lymphocytes
; IMMEDIATE SOURCE:
; LIBRARY: DM414 scFv antibodies obtained from
; LIBRARY: FUSE5 fusion phage construct
; CLONE: H18
; FEATURE:
; NAME/KEY: heavy chain
; US-08-983-607-20

```

```

Query Match 42.0%; Score 37; DB 4; Length 121;
Best Local Similarity 60.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 3; Indels 3; Gaps 0;

```

```

QY 2 WDTSALAPYL 11
Db 72 WDTSILTAYM 81

```

```

RESULT 33
US-08-233-146-1
; Sequence 1, Application US/08233146
; Patent No. 5648256
; GENERAL INFORMATION:
; APPLICANT: BEPPU, TERUHIKO
; APPLICANT: YAMADA, HIDEAKI
; APPLICANT: NAGASAWA, TORU
; APPLICANT: HORINOUCI, SUEHARU
; APPLICANT: NISHIYAMA, MAKOTO
; TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
; TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
; TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,146
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,746
; FILING DATE: 02-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. LESLIE
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7005-024-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas chlororaphis
; STRAIN: B23 (FERM BP-187)-
; US-08-463-470-1

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas chlororaphis
; STRAIN: B23 (FERM BP-187)
; US-08-233-146-1

```

```

Query Match 42.0%; Score 37; DB 1; Length 200;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 WDTSALAPYL 11
Db 156 WDTSAESRYL 165

```

```

RESULT 33
US-08-463-470-1
; Sequence 1, Application US/08463470
; Patent No. 5789211
; GENERAL INFORMATION:
; APPLICANT: BEPPU, TERUHIKO
; APPLICANT: YAMADA, HIDEAKI
; APPLICANT: NAGASAWA, TORU
; APPLICANT: HORINOUCI, SUEHARU
; APPLICANT: NISHIYAMA, MAKOTO
; TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
; TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
; TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,470
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,146
; FILING DATE: 22-APR-1994
; APPLICATION NUMBER: US 07/694,746
; FILING DATE: 02-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. LESLIE
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7005-024-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas chlororaphis
; STRAIN: B23 (FERM BP-187)-
; US-08-463-470-1

```

Query Match 42.0%; Score 37; DB 1; Length 200;  
Best Local Similarity 70.0%; Pred. No. 36;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WDTSAAPYL 11  
| | | | | : | |  
Db 156 WDTSAESRYL 165

RESULT 34

US-08-993-359-26  
; Sequence 26, Application US/08993359A  
; Patent No. 6039942  
; GENERAL INFORMATION:  
; APPLICANT: Lassen, Soren F.  
; APPLICANT: Bech, Lisbeth  
; APPLICANT: Ohmann, Anders  
; APPLICANT: Breinholt, Jens  
; APPLICANT: Fuglsang, Claus C.  
; APPLICANT: Ostergaard, Peter R.  
; TITLE OF INVENTION: Phytase Polypeptides  
; FILE REFERENCE: 5383.500-US  
; CURRENT APPLICATION NUMBER: US/08/993,359A  
; CURRENT FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 1480/96  
; EARLIER FILING DATE: 1996-12-20  
; EARLIER APPLICATION NUMBER: 1481/96  
; EARLIER FILING DATE: 1996-12-20  
; EARLIER APPLICATION NUMBER: 0301/97  
; EARLIER FILING DATE: 1997-03-18  
; EARLIER APPLICATION NUMBER: 0529/97  
; EARLIER FILING DATE: 1997-05-07  
; EARLIER APPLICATION NUMBER: 1388/97  
; EARLIER FILING DATE: 1997-12-01  
; EARLIER APPLICATION NUMBER: 60/046,082  
; EARLIER FILING DATE: 1997-05-09  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 26  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Paxillus involtus  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(19)  
US-08-993-359-26

Query Match 42.0%; Score 37; DB 3; Length 442;  
Best Local Similarity 54.5%; Pred. No. 91;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSAAPYL 12  
| | | | | : | |  
Db 367 WRTSSLVPFSG 377

RESULT 35

US-08-620-605D-2  
; Sequence 2, Application US/08620605D  
; Patent No. 5846811  
; GENERAL INFORMATION:  
; APPLICANT: SHIBATANI, TAKEJI  
; APPLICANT: AKATSUKA, HIROYUKI  
; APPLICANT: KAWAI, ERI  
; TITLE OF INVENTION: GENE PARTICIPATING IN THE MECHANISM OF  
; TITLE OF INVENTION: SECRETION OF ESTERASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH

STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/620,605D  
FILING DATE: 22-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 0020-3955  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 588 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-620-605D-2

Query Match 40.9%; Score 36; DB 2; Length 588;  
Best Local Similarity 45.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAAPYL 12  
| : | | | : |  
Db 399 WNKDELGPYIG 409

RESULT 36

US-09-005-232A-2  
; Sequence 2, Application US/09005232A  
; Patent No. 5922568  
; GENERAL INFORMATION:  
; APPLICANT: SHIBATANI, TAKEJI  
; APPLICANT: AKATSUKA, HIROYUKI  
; APPLICANT: KAWAI, ERI  
; TITLE OF INVENTION: GENE PARTICIPATING IN THE MECHANISM OF  
; TITLE OF INVENTION: SECRETION OF ESTERASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005,232A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ ID NO: 2:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 588 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-005-232A-2

Query Match      40.9%; Score 36; DB 2; Length 588;
Best Local Similarity 45.5%; Pred. NO. 1.9e+02;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      2 WDTSALAPYL 12
      | | | | |
Db      399 WNKDELGPYIG 409

RESULT 37
US-08-225-989-2
; Sequence 2, Application US/08225989
; Patent No. 5480981
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: NO. 5480981el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 892,459
; FILING DATE: 02-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 889,717
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2804-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-225-989-2

Query Match      40.9%; Score 36; DB 1; Length 595;
Best Local Similarity 57.1%; Pred. NO. 1.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      3 DTSALAPYLGTQEE 16
      | | | | |
Db      321 DTFEAPPLGTQPD 334

RESULT 38
US-08-570-923-2
; Sequence 2, Application US/08570923
; Patent No. 5677430
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: NO. 5677430el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,923
; FILING DATE: 12-DEC-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 892,459
; FILING DATE: 02-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 889,717
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2804-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-08-570-923-2

Query Match 40.9%; Score 36; DB 1; Length 595;  
Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16  
||: || ||||| :  
Db 321 DTTFEAPPLGTQPD 334

RESULT 39

US-08-580-014-2  
; Sequence 2, Application US/08580014  
; Patent No. 5753203  
; GENERAL INFORMATION:  
; APPLICANT: Goodwin, Raymond G.  
; APPLICANT: Smith, Craig A.  
; APPLICANT: Armitage, Richard J.  
; APPLICANT: Gruss, Hans-Jurgen  
; TITLE OF INVENTION: No. 5753203el Cytokine That Binds CD30  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/580,014  
FILING DATE: 20-DEC-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,989  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: US 07/966,775  
FILING DATE: 27-OCT-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 907,224  
FILING DATE: 01-JUL-1992  
APPLICATION NUMBER: US 899,660  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 892,459  
FILING DATE: 02-JUN-1992  
APPLICATION NUMBER: US 889,717  
FILING DATE: 26-MAY-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2804-E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-580-014-2

Query Match 40.9%; Score 36; DB 1; Length 595;  
Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16  
||: || ||||| :  
Db 321 DTTFEAPPLGTQPD 334

RESULT 40

US-08-232-087A-2  
; Sequence 2, Application US/08232087A  
; Patent No. 5866372  
; GENERAL INFORMATION:  
; APPLICANT: Stein, Harald  
; APPLICANT: D rkop, Horst  
; APPLICANT: Latza, Ute  
; TITLE OF INVENTION: Lymphoid CD30-Antigen  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
; STREET: 8110 Gatehouse Road, Suite 500 East  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22042

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,087A  
FILING DATE: 08-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 756-103P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-232-087A-2

Query Match 40.9%; Score 36; DB 2; Length 595;  
Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16  
||: || ||||| :  
Db 321 DTTFEAPPLGTQPD 334

RESULT 41

US-09-079-785-2  
; Sequence 2, Application US/09079785  
; Patent No. 6143869  
; GENERAL INFORMATION:  
; APPLICANT: Goodwin, Raymond G.  
; APPLICANT: Smith, Craig A.  
; APPLICANT: Armitage, Richard J.  
; APPLICANT: Gruss, Hans-Jurgen  
; TITLE OF INVENTION: No. 6143869el Cytokine That Binds CD30  
; NUMBER OF SEQUENCES: 23



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Kathryn A. Seese, Immunex Corporation  
;; STREET: 51 University Street  
;; CITY: Seattle  
;; STATE: Washington  
;; COUNTRY: USA  
;; ZIP: 98101  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: Apple Macintosh  
;; OPERATING SYSTEM: Apple 7.1  
;; SOFTWARE: Microsoft Word, Version 5.1a  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/079,785  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/225,989  
;; FILING DATE: 12 APRIL 1994  
;; APPLICATION NUMBER: US 07/966,775  
;; FILING DATE: 27-OCT-1992  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 907,224  
;; FILING DATE: 01-JUL-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 899,660  
;; FILING DATE: 15-JUN-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 892,459  
;; FILING DATE: 02-JUN-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 889,717  
;; FILING DATE: 26-MAY-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seese, Kathryn A.  
;; REGISTRATION NUMBER: 32,172  
;; REFERENCE/DOCKET NUMBER: 2804-E  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (206)587-0430  
;; TELEFAX: (206)233-0644  
;; TELEX: 756822  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 595 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-079-785-2

Query Match 40.9%; Score 36; DB 4; Length 595;  
Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16  
Db 321 DTTFEAPPLGTQPD 334  
RESULT 42  
US-08-815-175-3  
; Sequence 3, Application US/08815175  
; Patent No. 5856139  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL PROLINE-RICH ACIDIC PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive

;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: US  
;; ZIP: 94304  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSEQ Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/815,175  
;; FILING DATE: Filed Herewith  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Billings, Lucy J.  
;; REGISTRATION NUMBER: 36,749  
;; REFERENCE/DOCKET NUMBER: PF-0225 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-855-0555  
;; TELEFAX: 415-845-4166  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 149 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: GenBank  
;; CLONE: 899433  
US-08-815-175-3

Query Match 39.8%; Score 35; DB 2; Length 149;  
Best Local Similarity 75.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WDTSA LAP 9  
Db 45 WDTSA LAP 52

RESULT 43  
US-08-852-809-5  
; Sequence 5, Application US/08852809  
; Patent No. 5831052  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NEW HUMAN TRANSLOCATION ASSOCIATED  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,809  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0238 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1063411
;
US-08-852-809-5

Query Match 39.8%; Score 35; DB 2; Length 205;
Best Local Similarity 46.2%; Pred. NO. 84;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 4 TSALAPYLGTQEE 16
| :| | :| |
Db 61 TDSIGPYLYSKEE 73

RESULT 44
US-08-414-926A-22
; Sequence 22, Application US/08414926A
; Patent No. 5721354
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,926A
; FILING DATE: March 31, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.16
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..316
;
US-08-852-809-5

Query Match 39.8%; Score 35; DB 2; Length 205;
Best Local Similarity 46.2%; Pred. NO. 84;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 4 TSALAPYLGTQEE 16
| :| | :| |
Db 61 TDSIGPYLYSKEE 73

RESULT 44
US-08-414-926A-22
; Sequence 22, Application US/08414926A
; Patent No. 5721354
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,926A
; FILING DATE: March 31, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.16
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..316
;
US-08-414-926A-22

Query Match 39.8%; Score 35; DB 1; Length 316;
Best Local Similarity 37.5%; Pred. NO. 1.4e+02;
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 DWDTSALAPYLGTQEE 16
| | | :| :| |
Db 60 DWDSTMHCTPFWSTDLE 75

RESULT 45
US-08-926-922-22
; Sequence 22, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.16
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..316
; OTHER INFORMATION: /label= UL148
;
US-08-926-922-22

Query Match 39.8%; Score 35; DB 2; Length 316;
Best Local Similarity 37.5%; Pred. NO. 1.4e+02;
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 DWDTSALAPYLGTQEE 16
| | | :| :| |
Db 60 DWDSTMHCTPFWSTDLE 75

Search completed: June 28, 2001, 11:56:12
Job time: 193 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:59 ; Search time 42.81 Seconds  
(without alignments)  
28.470 Million cell updates/sec

Title: US-09-439-313-562  
Perfect score: 88  
Sequence: 1 DWDTSALAPYLGTQEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : PIR\_68:  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	51.1	685	2 G82022	methionine--trna 1
2	45	51.1	685	2 A81246	methionyl-trna syn
3	45	51.1	831	2 T00323	chitinase (EC 3.2.
4	45	51.1	847	2 F83517	probable penicilli
5	43	48.9	275	2 D83536	hypothetical prote
6	43	48.9	402	2 S74332	hypothetical prote
7	43	48.9	459	2 T35317	probable serine/th
8	43	48.9	626	2 E70747	hypothetical prote
9	42	47.7	207	2 S15205	superoxide dismuta
10	42	47.7	490	2 C81748	DNA gyrase, chain
11	42	47.7	643	2 S00480	potassium channel
12	42	47.7	656	2 JH0193	potassium channel
13	42	47.7	4485	2 T08044	dynein gamma heavy
14	41	46.6	217	2 E71098	probable aroM prot
15	41	46.6	233	1 DSBYN	superoxide dismuta
16	41	46.6	343	2 T31269	4-hydroxy-2-oxoval
17	41	46.6	415	2 C69682	gamma-glutamyl pho
18	41	46.6	427	2 D72528	probable L-asparag
19	40	45.5	207	2 S06599	superoxide dismuta
20	40	45.5	258	2 T31723	hypothetical prote
21	40	45.5	294	2 F83371	hypothetical prote
22	40	45.5	484	2 D65230	hypothetical 52.9
23	40	45.5	484	2 A86116	hypothetical prote
24	40	45.5	492	2 T38093	probable protein d
25	40	45.5	495	2 D86442	probable amino aci
26	40	45.5	732	2 C84487	hypothetical prote
27	40	45.5	957	2 E84547	probable disease r
28	40	45.5	1265	2 T51314	probable CO-induce
29	39.5	44.9	575	2 E84184	hypothetical prote

30	39	44.3	81	2 A70825	hypothetical prote
31	39	44.3	175	2 G70823	hypothetical prote
32	39	44.3	320	2 S73413	transcription anti
33	39	44.3	337	2 T32099	hypothetical prote
34	39	44.3	346	2 C71103	probable aspartoki
35	39	44.3	347	2 D84333	alcohol dehydrogen
36	39	44.3	387	2 S75981	magnesium/cobalt t
37	39	44.3	453	2 A65031	biphenyl dioxygena
38	39	44.3	453	2 G85898	hypothetical prote
39	39	44.3	453	2 A69494	Glu-trna amidotran
40	39	44.3	504	2 H69333	medium-chain acyl-
41	39	44.3	580	2 G83339	hypothetical prote
42	39	44.3	677	1 SYECMT	methionine--trna 1
43	39	44.3	677	2 C85839	methionine trna sy
44	39	44.3	677	2 C83210	methionyl-trna syn
45	39	44.3	682	2 I64113	methionine--trna 1
46	39	44.3	702	2 C82792	methionyl-trna syn
47	39	44.3	727	2 A56879	diacylglycerol kin
48	39	44.3	731	2 D82249	methionyl-trna syn
49	39	44.3	734	1 S09156	diacylglycerol kin
50	39	44.3	735	2 S12969	diacylglycerol kin
51	39	44.3	948	2 G83264	hypothetical prote
52	38	43.2	224	2 T04447	hypothetical prote
53	38	43.2	247	2 S74902	aquaporin 2 - Syne
54	38	43.2	256	2 T01189	anthocyanin biosyn
55	38	43.2	264	2 T03974	anthocyanin biosyn
56	38	43.2	266	2 T03972	anthocyanin biosyn
57	38	43.2	267	2 T01188	anthocyanin biosyn
58	38	43.2	268	2 T03729	anthocyanin biosyn
59	38	43.2	271	2 T03715	anthocyanin biosyn
60	38	43.2	277	2 F85530	carboxylesterase (
61	38	43.2	321	2 T38427	probable glucan 1,
62	38	43.2	541	2 T34701	hypothetical prote
63	38	43.2	544	2 B75379	CTP synthase - Dei
64	38	43.2	562	1 UKHUT	t-plasminogen acti
65	38	43.2	609	2 JQ0703	UDPglucose--starch

ALIGNMENTS

RESULT 1  
G82022 methionine--trna ligase (EC 6.1.1.10) NMA0275 [imported] - Neisseria meningitidis (st  
C;Species: Neisseria meningitidis  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C;Accession: G82022  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A;Reference number: A81775; MUID:20222556  
A;Accession: G82022  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-685 <PAR>  
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83583.1; PID:g737  
A;Experimental source: serogroup A, strain Z2491  
C;Genetics:  
A;Gene: metG; NMA0275  
C;Superfamily: methionine--trna ligase  
C;Keywords: ligase

Query Match 51.1%; Score 45; DB 2; Length 685;  
Best Local Similarity 66.7%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLG 12

Db 237 DWDISRDPYFG 248



RESULT 2  
 A:Title: Complete genome sequence of *Neisseria meningitidis* (strain MC58 serogroup)  
 A:Reference number: A81246  
 A:Accession: F83517  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-847 <STO>  
 A:Cross-references: GB:AE004535; GB:AE004091; NID:99946936; PIDN:AAG04421.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA1032

Query Match 51.1%; Score 45; DB 2; Length 685;  
 Best Local Similarity 66.7%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWTSALAPYLG 12  
 ||||| |||||  
 Db 237 DWDSRDAFYFG 248

RESULT 3  
 T00323  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A:Reference number: A81000; MUID:20175755  
 A:Accession: A81246  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-685 <TET>  
 A:Cross-references: GB:AE002361; GB:AE002098; NID:g7225245; PIDN:AAF40501.1; PID:g722524  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB0030  
 C:Superfamily: methionine--trRNA ligase

Query Match 51.1%; Score 45; DB 2; Length 685;  
 Best Local Similarity 66.7%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWTSALAPYLG 12  
 ||||| |||||  
 Db 237 DWDSRDAFYFG 248

RESULT 3  
 T00323  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A:Reference number: A81000; MUID:20175755  
 A:Accession: A81246  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-685 <TET>  
 A:Cross-references: GB:AE002361; GB:AE002098; NID:g7225245; PIDN:AAF40501.1; PID:g722524  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB0030  
 C:Superfamily: methionine--trRNA ligase

Query Match 51.1%; Score 45; DB 2; Length 831;  
 Best Local Similarity 64.3%; Pred. No. 16;  
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLG 15  
 ||||| |||||  
 Db 400 WDSSAKAPLYNPE 413

RESULT 4  
 F83517  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A:Reference number: A81000; MUID:20175755  
 A:Accession: A81246  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-685 <TET>  
 A:Cross-references: GB:AE002361; GB:AE002098; NID:g7225245; PIDN:AAF40501.1; PID:g722524  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB0030  
 C:Superfamily: methionine--trRNA ligase

Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: F83517  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-847 <STO>  
 A:Cross-references: GB:AE004535; GB:AE004091; NID:99946936; PIDN:AAG04421.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA1032

Query Match 51.1%; Score 45; DB 2; Length 847;  
 Best Local Similarity 52.9%; Pred. No. 16;  
 Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 DW--DTSALAPYLG 15  
 ||||| |||||  
 Db 721 EWQSDSSKNAPYLGAGE 737

RESULT 5  
 D83536  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: D83536  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-275 <STO>  
 A:Cross-references: GB:AE004522; GB:AE004091; NID:99946768; PIDN:AAG04267.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA0878

Query Match 48.9%; Score 43; DB 2; Length 275;  
 Best Local Similarity 66.7%; Pred. No. 10;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 SALAPYLG 16  
 ||||| |||||  
 Db 4 SAFAPWIGRQEE 15

RESULT 6  
 S74332  
 A:Title: Complete genome sequence of *Synechocystis* sp. (strain PCC 6803)  
 A:Reference number: S74332; MUID:97061201  
 A:Accession: S74332  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-402 <KAN>  
 A:Cross-references: EMBL:D64001; GB:AB001339; NID:gl001102; PIDN:BAAL0250.1; PID:d101  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 48.9%; Score 43; DB 2; Length 402;  
Best Local Similarity 53.8%; Pred. No. 16;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQ 14  
I: I: I I I I I  
Db 6 WENFAVAPYAGEQ 18

RESULT 7  
T35317  
probable serine/threonine protein kinase - Streptomyces coelicolor (fragment)  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35317  
R;Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z21575  
A:Accession: T35317  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-459 <OLI>  
A:Cross-references: EMBL:AL049863; PIDN:CAB42926.1; GSPDB:GN00070; SCOEDB:SC5H1.01  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC5H1.01

Query Match 48.9%; Score 43; DB 2; Length 459;  
Best Local Similarity 58.8%; Pred. No. 18;  
Matches 10; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 DWDTS----ALAPYLGT 13  
I I I I I I I I I I I I  
Db 437 DWDDSANERAAFLGT 453

RESULT 8  
E70747  
hypothetical protein Rv2917 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: E70747  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: E70747  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-626 <COL>  
A:Cross-references: GB:274697; GB:AL123456; NID:g3261602; PIDN:CAA98979.1; PID:e326856;  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv2917

Query Match 48.9%; Score 43; DB 2; Length 626;  
Best Local Similarity 61.5%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQ 14  
I I I I I I I I I I I I  
Db 68 WQRALVQYLGTQ 80

RESULT 9  
SI5205  
superoxide dismutase (EC 1.15.1.1) (Fe) [validated] - Mycobacterium tuberculosis

C:Species: Mycobacterium tuberculosis  
C:Date: 21-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 20-Jun-2000  
C:Accession: SI5205; C70654; SI0908  
R;Zhang, Y.; Lathigra, R.; Garbe, T.; Catty, D.; Young, D.  
Mol. Microbiol. 5, 381-391, 1991  
A:Title: Genetic analysis of superoxide dismutase, the 23 kilodalton antigen of Mycob  
A:Reference number: SI5205; MUID:91251768  
A:Accession: SI5205  
A:Molecule type: DNA  
A:Residues: 1-207 <ZHA>  
A:Cross-references: EMBL:X52861; NID:g794079; PIDN:CAA37042.1; PID:g581379  
A:Experimental source: strain H37RV  
A:Note: the authors translated the codon GAA for residue 3 as Gln  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70654  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-207 <COL>  
A:Cross-references: GB:Z83864; GB:AL123456; NID:g3261687; PIDN:CAB06220.1; PID:gl7811  
A:Experimental source: strain H37RV  
R;Cooper, J.B.; McIntyre, K.; Wood, S.P.; Zhang, Y.; Young, D.  
submitted to the Brookhaven Protein Data Bank, September 1994  
A:Reference number: A52807; PDB:1IDS  
A:Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 2-199  
R;Cooper, J.B.; Driessen, H.P.C.; Wood, S.P.; Zhang, Y.; Young, D.  
J. Mol. Biol. 235, 1156-1158, 1994  
A:Title: Crystallisation and preliminary X-ray analysis of the iron-dependent superox  
A:Reference number: A59029; MUID:94118350  
A:Contents: annotation; X-ray crystallography  
C:Genetics:  
A:Gene: soda  
A:Start codon: GTG  
C:Function:  
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg  
C:Superfamily: superoxide dismutase (Mn)  
C:Keywords: iron; metalloprotein; oxidoreductase; tetramer  
F;28,76,160,164/Binding site: iron (His, His, Asp, His) #status experimental

Query Match 47.7%; Score 42; DB 2; Length 207;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQ 14  
I I I I I I I I I I I  
Db 10 DWDYGALEPHISGQ 23

RESULT 10  
C81748  
DNA gyrase, chain A TC0031 [imported] - Chlamydia muridarum (strain Nigg)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Jun-2000  
C:Accession: C81748  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
; C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: C81748  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-490 <TET>  
A:Cross-references: GB:AE002271; GB:AE002160; NID:g7190063; PIDN:AAF38922.1; PID:g719  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:  
A:Gene: TC0031



C;Species: Chlamydomonas reinhardtii  
C;Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 02-Feb-2001  
C;Accession: T08044  
R;Wilkinson, C.G.; King, S.M.; Witman, G.B.  
J. Cell Sci. 107, 497-506, 1994  
A;Title: Molecular analysis of the gamma heavy chain of Chlamydomonas flagellar outer-antennal  
A;Reference number: Z16311; MUID:94274766  
A;Accession: T08044  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-4485 <WIL>  
A;Cross-references: EMBL:U15303; NID:g557715; PIDN:AAA50455.1; PID:g557716  
A;Experimental source: strain 1132D  
C;Superfamily: dynein heavy chain, ciliary  
C;Keywords: ATP; hydrolase; microtubule binding; nucleotide binding; P-loop  
F;1819-1826/Region: nucleotide-binding motif A (P-loop)

Query Match 47.7%; Score 42; DB 2; Length 4485;  
Best Local Similarity 46.7%; Pred. No. 3.4e+02;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQE 15  
||: ||: ||: ||  
Db 1188 DWEANAMVPGLDPOE 1202

RESULT 14  
E71098  
probable aroM protein - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
C;Accession: E71098  
R;Kawarabayashi, Y.; Savada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A;Reference number: A71000; MUID:98344137  
A;Accession: E71098  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-217 <KAW>  
A;Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30147.1; PID:g3257464  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH1049  
C;Superfamily: aroM protein

Query Match 46.6%; Score 41; DB 2; Length 217;  
Best Local Similarity 54.5%; Pred. No. 17;  
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALAPYLGTQEE 16  
::||: ||: ||  
Db 154 SVSPYVGTEEE 164

RESULT 15  
DSBYN  
superoxide dismutase (EC 1.15.1.1) (Mn) precursor [validated] - yeast (Saccharomyces cer  
N;Alternate names: protein YHR008c  
C;Species: Saccharomyces cerevisiae  
C;Date: 19-Feb-1984 #sequence\_revision 17-Mar-1987 #text\_change 23-Mar-2001  
C;Accession: A00521; S46785; A90766  
R;Marres, C.A.M.; Van Loon, A.P.G.M.; Oudshoorn, P.; Van Steeg, H.; Grivell, L.A.; Slate  
Eur. J. Biochem. 147, 153-161, 1985  
A;Title: Nucleotide sequence analysis of the nuclear gene coding for manganese superoxid  
A;Reference number: A91141; MUID:85127011  
A;Accession: A00521  
A;Molecule type: DNA  
A;Residues: 1-233 <MAR>

A;Cross-references: EMBL:X02156; NID:g4513; PIDN:CAA26092.1; PID:g4514  
R;Du, Z.  
submitted to the EMBL Data Library, June 1994  
A;Description: The sequence of S. cerevisiae cosmid L2825.  
A;Reference number: S46774  
A;Accession: S46785  
A;Molecule type: DNA  
A;Residues: 1-233 <DUZ>  
A;Cross-references: EMBL:U10400; NID:g500701; PIDN:AAB68939.1; PID:g500704; GSPDB:GNO  
R;Ditlow, C.; Johansen, J.T.; Martin, B.M.; Svendsen, I.  
Carlsberg Res. Commun. 47, 81-91, 1982  
A;Title: The complete amino acid sequence of manganese-superoxide dismutase from Sacc  
A;Reference number: A90766  
A;Accession: A90766  
A;Molecule type: protein  
A;Residues: 27-229 <DIT>  
C;Genetics:  
A;Gene: SGD:SOD2; MIPS:YHR008c  
A;Cross-references: SGD:S0001050; MIPS:YHR008c  
A;Map position: 8R  
A;Genome: nuclear  
C;Complex: homotetramer  
C;Function:  
A;Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg  
C;Superfamily: superoxide dismutase (Mn)  
C;Keywords: homotetramer; manganese; metalloprotein; mitochondrial matrix; mitochondr  
F;1-26/Domain: transit peptide (mitochondrion) #status predicted <TNP>  
F;27-229/Product: superoxide dismutase (Mn) #status experimental <MAT>  
F;52,107,194,198/Binding site: manganese (His, His, Asp, His) #status predicted

Query Match 46.6%; Score 41; DB 1; Length 233;  
Best Local Similarity 53.8%; Pred. No. 19;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTsALAPYLGTQ 14  
|| || || ||: |  
Db 35 WDFGALEPYISGQ 47

RESULT 16  
T31269  
4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.-) - Sphingomonas aromaticivorans plasmid  
C;Species: Sphingomonas aromaticivorans  
C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000  
C;Accession: T31269  
R;Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.  
submitted to the EMBL Data Library, July 1998  
A;Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas arom  
A;Reference number: 220992  
A;Accession: T31269  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-343 <ROM>  
A;Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378410; PIDN:AAD03993.1  
C;Genetics:  
A;Gene: xylK  
A;Genome: plasmid pNL1  
C;Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 46.6%; Score 41; DB 2; Length 343;  
Best Local Similarity 33.3%; Pred. No. 29;  
Matches 10; Conservative 2; Mismatches 4; Indels 14; Gaps 1;

QY 1 DWD-----TSALAPYLGTQEE 16  
||| | : | | | | |  
Db 70 DWDWIEAAADVIKNAVLTTLLVPGIGTAEE 99

RESULT 17  
C69682  
gamma-glutamyl phosphate reductase proA - Bacillus subtilis



C;Species: Bacillus subtilis  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C;Accession: C69682  
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, K.; Tosato, V.; Uchiyama  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A;Reference number: A69580; MUID:98044033  
A;Accession: C69682  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-415 <KUN>  
A;Cross-references: GB:Z99110; GB:AL009126; NID:g26333472; PIDN:CAB13170.1; PID:g26333667  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: proA  
C;Superfamily: glutamate-5-semialdehyde dehydrogenase

Query Match	46.6%	Score 41;	DB 2;	Length 415;
Best Local Similarity	50.0%	Pred. No. 36;		
Matches	7;	Conservative	2;	Mismatches 5; Indels 0; Gaps 0;

QY	1	DWDTSALAPYLGTQ 14
Db	306	DWETELAPVLSVK 319

RESULT 18  
D72528  
probable L-asparaginase APE2200 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: D72528  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A;Reference number: A72450; MUID:99310339  
A;Accession: D72528  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-427 <KAW>  
A;Cross-references: DDBJ:AP000063; NID:g5105654; PIDN:BAA81212.1; PID:g5105900  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE2200

Query Match	46.6%	Score 41;	DB 2;	Length 427;
Best Local Similarity	46.7%	Pred. No. 37;		
Matches	7;	Conservative	3;	Mismatches 5; Indels 0; Gaps 0;

QY	1	DWDTSALAPYLGTQE 15
Db	90	DYETGAVKPYLDASE 104

RESULT 19  
S06599  
superoxide dismutase (EC 1.15.1.1) (Mn) - Mycobacterium leprae  
C;Species: Mycobacterium leprae

C;Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 18-Jun-1999  
C;Accession: S06599; A37212  
R;Thangaraj, H.S.; Lamb, F.I.; Davis, E.O.; Colston, M.J.  
Nucleic Acids Res. 17, 8378, 1989  
A;Title: Nucleotide and deduced amino acid sequence of Mycobacterium leprae manganese  
A;Reference number: S06599; MUID:90045970  
A;Accession: S06599  
A;Molecule type: DNA  
A;Residues: 1-207 <THA>  
A;Cross-references: EMBL:X16453; NID:g44405; PIDN:CAA34472.1; PID:g581343  
R;Thangaraj, H.S.; Lamb, F.I.; Davis, E.O.; Jenner, P.J.; Jeyakumar, L.H.; Colston, M  
Infect. Immun. 58, 1937-1942, 1990  
A;Title: Identification, sequencing, and expression of Mycobacterium leprae superoxid  
A;Reference number: A37212; MUID:90256282  
A;Accession: A37212  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-207 <TH2>  
A;Cross-references: GB:X16453  
C;Genetics:  
A;Start codon: GTG  
C;Function:  
A;Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg  
C;Superfamily: superoxide dismutase (Mn)  
C;Keywords: manganese; metalloprotein; oxidoreductase  
F;28,76,160,164/Binding site: manganese (His, His, Asp, His) #status predicted

Query Match	45.5%	Score 40;	DB 2;	Length 207;
Best Local Similarity	54.5%	Pred. No. 24;		
Matches	6;	Conservative	3;	Mismatches 2; Indels 0; Gaps 0;

QY	1	DWDTSALAPYL 11
Db	10	DWDYAALEPHI 20

RESULT 20  
T31723  
hypothetical protein F40A3.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 04-Mar-2000  
C;Accession: T31723  
R;Geisel, C.; Bradshaw, H.; Keppler, D.  
submitted to the EMBL Data Library, July 1997  
A;Description: The sequence of C. elegans cosmid F40A3.  
A;Reference number: Z21074  
A;Accession: T31723  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-258 <GEI>  
A;Cross-references: EMBL:AF016423; PIDN:AAB65325.1; GSPDB:GN00023; CESP:F40A3.6  
A;Experimental source: strain Bristol N2; clone F40A3  
C;Genetics:  
A;Gene: CESP:F40A3.6  
A;Map position: 5  
A;Introns: 53/3; 109/3; 175/3; 204/2  
C;Superfamily: Caenorhabditis elegans hypothetical protein F40A3.6

Query Match	45.5%	Score 40;	DB 2;	Length 258;
Best Local Similarity	63.6%	Pred. No. 31;		
Matches	7;	Conservative	3;	Mismatches 1; Indels 0; Gaps 0;

QY	1	DWDTSALAPYL 11
Db	201	DFDSSPLSPYL 211

RESULT 21  
F83371  
hypothetical protein PA2201 [imported] - Pseudomonas aeruginosa (strain PA01)  
C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: F83371  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A;Reference number: A82950; MUID:20437337  
A;Accession: F83371  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-294 <STO>  
A;Cross-references: GB:AE004646; GB:AE004091; NID:g9948213; PIDN:AAG05589.1; GSPDB:GN00145  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA2201

Query Match 45.5%; Score 40; DB 2; Length 294;  
Best Local Similarity 66.7%; Pred. No. 36;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQ 14  
| ||||| :|  
Db 218 DAQALAPYLQSQ 229

RESULT 22  
D65230  
hypothetical 52.9 kD protein in aidB-rpsF intergenic region - *Escherichia coli* (strain K12)  
N;Alternate names: hypothetical protein o488  
C;Species: *Escherichia coli*  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C;Accession: D65230; S56418  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of *Escherichia coli* K-12.  
A;Reference number: A64720; MUID:97426617  
A;Accession: D65230  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-484 <BLAT>  
A;Cross-references: GB:AE000491; GB:U00096; NID:g2367357; PIDN:AAC77150.1; PID:g2367358;  
A;Experimental source: strain K-12, substrain MGL655  
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23, 2105-2119, 1995  
A;Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.1 to 100.0%  
A;Reference number: S56314; MUID:95334362  
A;Accession: S56418  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-469, 'AQKMKNNQNSLLNKEF' <BUR>  
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97089.1; PID:g537034  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
C;Genetics:  
A;Gene: yjfs  
A;Start codon: GTG

Query Match 45.5%; Score 40; DB 2; Length 484;  
Best Local Similarity 46.7%; Pred. No. 62;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16  
| :||| :||  
Db 220 WIAYKVAPFLGKKEE 234

RESULT 23  
A86116  
hypothetical protein sgat [imported] - *Escherichia coli* (strain O157:H7)  
C;Species: *Escherichia coli*

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: A86116  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, N.; Lory, S.; Olson, M.V.  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: A86116  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-484 <STO>  
A;Cross-references: GB:AE005174; NID:g12519184; PIDN:AAG59389.1; GSPDB:GN00145; UWGP:  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: sgat

Query Match 45.5%; Score 40; DB 2; Length 484;  
Best Local Similarity 46.7%; Pred. No. 62;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16  
| :||| :||  
Db 220 WIAYKVAPFLGKKEE 234

RESULT 24  
T38093  
probable protein disulfide-isomerase (EC 5.3.4.1) SPAC1F5.02 [similarity] - fission yeast  
C;Species: *Schizosaccharomyces pombe*  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 15-Sep-2000  
C;Accession: T38093  
R;Gentles, S.; Odell, C.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, November 1995  
A;Reference number: Z21768  
A;Accession: T38093  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-492 <GEN>  
A;Cross-references: EMBL:Z68136; PIDN:CAA92230.1; GSPDB:GN00066; SPDB:SPAC1F5.02  
A;Experimental source: strain 972h-; cosmid c1F5  
C;Genetics:  
A;Gene: SPDB:SPAC1F5.02  
A;Map position: 1  
C;Superfamily: protein disulfide-isomerase; thioredoxin homology  
C;Keywords: intramolecular oxidoreductase; isomerase

Query Match 45.5%; Score 40; DB 2; Length 492;  
Best Local Similarity 41.7%; Pred. No. 64;  
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTALAPYLIG 12  
||| :|| :||  
Db 215 DWDPASIAADFIG 226

RESULT 25  
D86442  
probable amino acid permease [imported] - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: D86442  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, U.; Ben-Zur, N.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, D.; Drenth, C.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia, R.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719  
A:Accession: D86442  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-495 <STO>  
A:Cross-references: GB:AE005172; NID:g11136719; PIDN:AAG31300.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 45.5%; Score 40; DB 2; Length 495;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQE 15  
|||:|:|:  
Db 116 WVSSALGPFWGFQ 129

RESULT 26  
C84487  
hypothetical protein At2g07650 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: C84487  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-732 <STO>  
A:Cross-references: GB:AE002093; NID:g5001454; PIDN:AAD37021.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g07650  
A:Map position: 2

Query Match 45.5%; Score 40; DB 2; Length 732;  
Best Local Similarity 41.7%; Pred. No. 99;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGT 13  
||:|:|:  
Db 500 WDMATAPYISS 511

RESULT 27  
E84547  
probable disease resistance protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84547  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: E84547  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-957 <STO>  
A:Cross-references: GB:AE002093; NID:g6598361; PIDN:AAF18599.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g17050  
A:Map position: 2

Query Match 45.5%; Score 40; DB 2; Length 957;

Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQE 15  
|||:|:|:  
Db 335 WDTSLIQLLEDEE 348

RESULT 28  
T51314  
probable CO-induced hydrogenase complex chain COOM [imported] - Rhodospirillum rubrum  
C:Species: Rhodospirillum rubrum  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C:Accession: T51314  
R:Kerby, R.L.  
submitted to the EMBL Data Library, July 1996  
A:Reference number: Z25372  
A:Accession: T51314  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1265 <KER>  
A:Cross-references: EMBL:U65510; PIDN:AAC45116.1  
A:Experimental source: strain UR1  
C:Genetics:  
A:Gene: COOM

Query Match 45.5%; Score 40; DB 2; Length 1265;  
Best Local Similarity 54.5%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGT 12  
||:|:|:  
Db 229 WDAVLLPVG 239

RESULT 29  
E84184  
hypothetical protein Vng0243c [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84184  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: E84184  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-575 <STO>  
A:Cross-references: GB:AE004437; NID:g10579883; PIDN:AAG18841.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG0243C

Query Match 44.9%; Score 39.5; DB 2; Length 575;  
Best Local Similarity 50.0%; Pred. No. 92;  
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 1 DWDTSAALAPYLGTQEE 16  
||:|:|:  
Db 338 DWEEAAMA-YLGAPTE 352

RESULT 30  
A70825  
hypothetical protein Rv0750 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70825

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession:..A70825  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-81 <COL>  
A;Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17517.1; PID:e125328  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv0750

Query Match 44.3%; Score 39; DB 2; Length 81;  
Best Local Similarity 70.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 2 WDTSLAPYL 11  
| | | | |  
Db 60 WYTQGLAPYL 69

RESULT 31  
G70823  
hypothetical protein Rv0740 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: G70823  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: G70823  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-175 <COL>  
A;Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17507.1; PID:e125327  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv0740

Query Match 44.3%; Score 39; DB 2; Length 175;  
Best Local Similarity 70.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 2 WDTSLAPYL 11  
| | | | |  
Db 154 WYTQGLAPYL 163

RESULT 32  
S73413  
transcription antitermination factor nusG - Mycoplasma pneumoniae (strain ATCC 29342)  
N;Alternate names: hypothetical protein D09\_orf320  
C;Species: Mycoplasma pneumoniae  
A;Variety: ATCC 29342  
C;Date: 26-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999  
C;Accession: S73413  
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R. Nucleic Acids Res. 24, 4420-4449, 1996  
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
A;Reference number: S73327; MUID:97105885  
A;Accession: S73413  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA

A;Residues: 1-320 <HIM>  
A;Cross-references: EMBL:AE000011; GB:U00089; NID:g1673740; PIDN:AAB95735.1; PID:g167  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996  
C;Genetics:  
A;Gene: nusG  
A;Genetic code: SGC3

Query Match 44.3%; Score 39; DB 2; Length 320;  
Best Local Similarity 61.5%; Pred. No. 58;  
Matches 8; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

QY 3 DTSALAPYLGTQE 15  
| | | | | | |  
Db 234 DPSVLIPYLGYE 246

RESULT 33  
T32099  
hypothetical protein F45C12.8 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Aug-2000  
C;Accession: T32099  
R;Johnson, D.; Kramer, J.; Keppler, D.  
submitted to the EMBL Data Library, July 1997  
A;Description: The sequence of C. elegans cosmid F45C12.  
A;Reference number: Z21123  
A;Accession: T32099  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-337 <JOH>  
A;Cross-references: EMBL:AF016684; PIDN:AAB66207.1; GSPDB:GN00020; CESP:F45C12.8  
A;Experimental source: strain Bristol N2; clone F45C12  
C;Genetics:  
A;Gene: CESP:F45C12.8  
A;Map position: 2  
A;Introns: 19/2; 50/2; 78/2; 269/3  
C;Superfamily: Caenorhabditis elegans hypothetical protein C36C9.3

Query Match 44.3%; Score 39; DB 2; Length 337;  
Best Local Similarity 50.0%; Pred. No. 62;  
Matches 8; Conservative 2; Mismatches 6; Indels 6; Gaps 0;

QY 1 DWTSLAPYLGTQEE 16  
| : | | | | | | | | |  
Db 286 DMNTSGLKSVFGTQDE 301

RESULT 34  
C71103  
probable aspartokinase - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
C;Accession: C71103  
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A;Reference number: A71000; MUID:98344137  
A;Accession: C71103  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-346 <KAW>  
A;Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30185.1; PID:d1031128; PID:g32  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa  
C;Genetics:  
A;Gene: PH1086

Query Match 44.3%; Score 39; DB 2; Length 346;  
Best Local Similarity 77.8%; Pred. No. 64;



Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 DTSALAPYL 11  
 Db 80 DTSSLSPYL 88

RESULT 35  
 D84333  
 alcohol dehydrogenase [imported] - Halobacterium sp. NRC-1  
 C;Species: Halobacterium sp. NRC-1  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C;Accession: D84333  
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo, Jung, K.H.; Alam, M.; Freitas, T.  
 proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A;Title: Genome sequence of Halobacterium species NRC-1.  
 A;Reference number: A84160; MUID:20504483  
 A;Accession: D84333  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-347 <STO>  
 A;Cross-references: GB:AE004437; NID:gl0581268; PIDN:AAG20032.1; GSPDB:GN00138  
 C;Genetics:  
 A;Gene: adh4  
 C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 44.3%; Score 39; DB 2; Length 347;  
 Best Local Similarity 66.7%; Pred. No. 64;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DWDTALAP 9  
 Db 140 DWETAAAP 148

RESULT 36  
 S75981  
 magnesium/cobalt transport protein sll0507 - Synechocystis sp. (strain PCC 6803)  
 N;Alternate names: protein sll0507  
 C;Species: Synechocystis sp.  
 A;Variety: PCC 6803  
 C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C;Accession: S75981  
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 s.  
 A;Reference number: S74322; MUID:97061201  
 A;Accession: S75981  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-387 <KAN>  
 A;Cross-references: EMBL:D64006; GB:AB001339; NID:gl001291; PIDN:BAAL0828.1; PID:gl00134  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C;Superfamily: magnesium and cobalt transport protein

Query Match 44.3%; Score 39; DB 2; Length 387;  
 Best Local Similarity 61.5%; Pred. No. 72;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTALAPYLGT 13  
 Db 69 DISPALRPYLGT 81

RESULT 37  
 A65031

biphenyl dioxygenase (EC 1.14.-.-) terminal oxygenase component large chain - Escheri  
 C;Species: Escherichia coli  
 C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 20-Aug-1999  
 C;Accession: A65031; S49292  
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A;Title: The complete genome sequence of Escherichia coli K-12.  
 A;Reference number: A64720; MUID:97426617  
 A;Accession: A65031  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-453 <BLAT>  
 A;Cross-references: GB:AE000340; GB:U00096; NID:gl788883; PIDN:AAC75591.1; PID:gl78888  
 A;Experimental source: strain K-12, substrain MGL1655  
 R;Turlin, E.; Gasser, F.; Biville, F.  
 submitted to the EMBL Data Library, September 1994  
 A;Description: Cloning and sequencing of an E. coli gene homologous to dioxygenase of  
 A;Reference number: S49292  
 A;Accession: S49292  
 A;Molecule type: DNA  
 A;Residues: 1-19,'A',21-383,'ATAPATANCYWKW',397-398,451,'R',453,'SAATTAFLALLTISFQKLPL  
 A;Cross-references: EMBL:237966; NID:g550595; PIDN:CAA86018.1; PID:g550596  
 C;Superfamily: toluene dioxygenase terminal oxygenase component large chain; Rieske [  
 C;Keywords: 2Fe-2S; metalloprotein; oxidoreductase; Rieske iron-sulfur protein  
 F;75-123/Domain: Rieske [2Fe-2S] homology <RSK>  
 F;85,87,105,108/Binding site: 2Fe-2S cluster (Cys, His, Cys, His) (covalent) #status

Query Match 44.3%; Score 39; DB 2; Length 453;  
 Best Local Similarity 64.3%; Pred. No. 86;  
 Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 DWDTSA--LAPYLG 12  
 Db 155 NWDTSAPGLRDYLG 168

RESULT 38  
 G83898  
 hypothetical protein hcaA1 [imported] - Escherichia coli (strain Ol57:H7)  
 C;Species: Escherichia coli  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
 C;Accession: G85898  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: G85898  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-453 <STO>  
 A;Cross-references: GB:AE005174; NID:gl2516944; PIDN:AAG57651.1; GSPDB:GN00145; UWGP:  
 A;Experimental source: strain Ol57:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: hcaA1  
 C;Superfamily: toluene dioxygenase terminal oxygenase component large chain; Rieske [

Query Match 44.3%; Score 39; DB 2; Length 453;  
 Best Local Similarity 64.3%; Pred. No. 86;  
 Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 DWDTSA--LAPYLG 12  
 Db 155 NWDTSAPGLRDYLG 168

RESULT 39  
 A69494  
 Glu-tRNA amidotransferase, subunit A (gata-1) homolog - Archaeoglobus fulgidus  
 C;Species: Archaeoglobus fulgidus

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Jun-1999  
C;Accession: A69494  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A;Reference number: A69250; MUID:98049343  
A;Accession: A69494  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-453 <KLE>  
A;Cross-references: GB:AE000968; GB:AE000782; NID:g2689291; PIDN:AAB89301.1; PID:g264858  
C;Superfamily: indoleacetamide hydrolase

Query Match 44.3%; Score 39; DB 2; Length 453;  
Best Local Similarity 50.0%; Pred. No. 86;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL G 12  
:|: | |||  
Db 313 EWEKVAFP P YLG 324

RESULT 40  
H69333  
medium-chain acyl-CoA ligase (alkK-3) homolog - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 18-Aug-2000  
C;Accession: H69333  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A;Reference number: A69250; MUID:98049343  
A;Accession: H69333  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-504 <KLE>  
A;Cross-references: GB:AE001058; GB:AE000782; NID:g2689381; PIDN:AAB90569.1; PID:g264994  
C;Superfamily: probable acyl-CoA ligase medium chain; acetate--CoA ligase homology  
F;55-497/Domain: acetate--CoA ligase homology <ACL>

Query Match 44.3%; Score 39; DB 2; Length 504;  
Best Local Similarity 54.5%; Pred. No. 96;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDT SALAPYL G 12  
| | :|||:|  
Db 212 WGT VFIAPYMG 222

RESULT 41  
G83339  
hypothetical protein PA2448 [imported] - Pseudomonas aeruginosa (strain PAO1)  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: G83339  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho  
A;Reference number: A82950; MUID:20437337  
A;Accession: G83339  
A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-580 <STO>  
A;Cross-references: GB:AE004672; GB:AE004091; NID:g9948492; PIDN:AAG05836.1; GSPDB:GN  
A;Experimental source: strain PAO1  
C;Genetics:  
A;Gene: PA2448

Query Match 44.3%; Score 39; DB 2; Length 580;  
Best Local Similarity 40.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQE 15  
|| |::|: |:  
Db 480 DWPVSSLS PWKAIQQ 494

RESULT 42  
SYECMT  
methionine--tRNA ligase (EC 6.1.1.10) [validated] - Escherichia coli  
N;Alternate names: methionyl-tRNA synthetase  
C;Species: Escherichia coli  
C;Date: 13-Jun-1983 #sequence\_revision 30-Jun-1992 #text\_change 26-May-2000  
C;Accession: S14427; A91797; A91123; S11949; A64979; A01187; A35821  
R;Dardel, F.  
submitted to the EMBL Data Library, October 1990  
A;Reference number: S14427  
A;Accession: S14427  
A;Molecule type: DNA  
A;Residues: 1-677 <DAR1>  
A;Cross-references: EMBL:X55791; NID:g42015; PIDN:CAA39315.1; PID:g42016  
R;Dardel, F.; Fayat, G.; Blanquet, S.  
J. Bacteriol. 160, 1115-1122, 1984  
A;Title: Molecular cloning and primary structure of the Escherichia coli methionyl-tR  
A;Reference number: A91797; MUID:85054627  
A;Accession: A91797  
A;Molecule type: DNA  
A;Residues: 2-677 <DAR2>  
A;Cross-references: GB:K02671; NID:g146828; PIDN:AAA24161.1; PID:g146829; GB:J01649;  
R;Barker, D.G.; Ebel, J.P.; Jakes, R.; Bruton, C.J.  
Eur. J. Biochem. 127, 449-457, 1982  
A;Title: Methionyl-tRNA synthetase from E. coli: primary structure of the active crys  
A;Reference number: A91123; MUID:83079258  
A;Accession: A91123  
A;Molecule type: DNA  
A;Residues: 2-15, 'V', 17-149, 'L', 151-435, 'A', 437-564 <BAR>  
A;Cross-references: GB:K02671; GB:J01649; GB:J01650  
A;Note: most of this sequence was confirmed by protein sequencing  
R;Dardel, F.; Panvert, M.; Fayat, G.  
Mol. Gen. Genet. 223, 121-133, 1990  
A;Title: Transcription and regulation of expression of the Escherichia coli methionyl  
A;Reference number: S11948; MUID:91080852  
A;Accession: S11949  
A;Molecule type: DNA  
A;Residues: 1-51;661-677 <DAR>  
A;Cross-references: EMBL:X55791  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617  
A;Accession: A64979  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-677 <BLAT>  
A;Cross-references: GB:AE000300; GB:J000096; NID:g1788425; PIDN:AAC75175.1; PID:g17884  
A;Experimental source: strain K-12, substrain MG1655  
R;Zelwer, C.; Risler, J.L.; Brunie, S.  
J. Mol. Biol. 155, 63-81, 1982  
A;Title: Crystal structure of Escherichia coli methionyl-tRNA synthetase at 2.5 A res  
A;Reference number: A92880; MUID:82192427  
A;Contents: annotation; x-ray crystallography, 2.5 angstroms  
R;Hountondji, C.; Schmitter, J.M.; Beauvallet, C.; Blanquet, S.

Biochemistry 29, 8190-8198, 1990  
A;Title: Mapping of the active site of Escherichia coli methionyl-tRNA synthetase: identification of the 3'-acceptor end.  
A;Reference number: A35821; MUID:91084494  
A;Contents: annotation; active site  
R;Fourmy, D.; Dardel, F.  
Submitted to: The Brookhaven Protein Data Bank, November 1992  
A;Reference number: A51298; PDB:1MEA  
A;Contents: annotation; conformation by (1)H-NMR, residues 'GS',139-164  
R;Fourmy, D.; Dardel, F.; Blanquet, S.  
J. Mol. Biol. 231, 1078-1089, 1993  
A;Title: Methionyl-tRNA synthetase zinc binding domain. Three-dimensional structure and A;Reference number: A58691; MUID:93294859  
A;Contents: annotation; conformation by (1)H-NMR  
R;Brunie, S.; Zeller, C.; Risler, J.L.  
J. Mol. Biol. 216, 411-424, 1990  
A;Title: Crystallographic study at 2.5 Angstroms resolution of the interaction of methionyl-tRNA synthetase with methionyl-tRNA synthetase  
A;Reference number: A58692; MUID:91073404  
A;Contents: annotation; X-ray crystallography, 2.5 angstroms  
C;Genetics:  
A;Gene: metG  
A;Map position: 46 min  
C;Function:  
A;Description: EC 6.1.1.10 [validated; MUID:83079258]; catalyzes the ligation of methionine to tRNA  
A;Pathway: protein biosynthesis  
C;Superfamily: methionine--tRNA ligase  
C;Keywords: aminoacyl-tRNA synthetase; ATP; homodimer; ligase; metalloprotein; protein binding  
F;2-677/Product: methionine--tRNA ligase #status predicted <MAT>  
F;62-66,137-157,335-341,434-438/Region: tRNA 3'-acceptor end binding  
F;452-468/Region: anticodon recognition  
F;16,22,53,336/Active site: Tyr, His, Asp, Lys #status predicted  
F;146,149,159,162/Binding site: zinc (Cys) #status experimental

Query Match 44.3%; Score 39; DB 1; Length 677;  
Best Local Similarity 63.6%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLG 12  
|||  
Db 230 WDISRDAPYFG 240

RESULT 43  
C85839  
methionine tRNA synthetase [imported] - Escherichia coli (strain O157:H7)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: C85839  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: C85839  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-677 <STO>  
A;Cross-references: GB:AE005174; NID:g12516327; PIDN:AAG57175.1; GSPDB:GN00145; UWGP:232  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: metG  
C;Superfamily: methionine--tRNA ligase

Query Match 44.3%; Score 39; DB 2; Length 677;  
Best Local Similarity 63.6%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLG 12  
|||  
Db 230 WDISRDAPYFG 240

RESULT 44

C83210  
methionyl-tRNA synthetase PA3482 [imported] - Pseudomonas aeruginosa (strain PA01)  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: C83210  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A;Reference number: A82950; MUID:20437337  
A;Accession: C83210  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-677 <STO>  
A;Cross-references: GB:AE004769; GB:AE004091; NID:g9949624; PIDN:AAG06870.1; GSPDB:GN  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: metG; PA3482  
C;Superfamily: methionine--tRNA ligase

Query Match 44.3%; Score 39; DB 2; Length 677;  
Best Local Similarity 63.6%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLG 12  
|||  
Db 229 WDISRDAPYFG 239

RESULT 45  
I64113  
methionine--tRNA ligase (EC 6.1.1.10) - Haemophilus influenzae (strain Rd KW20)  
N;Alternate names: methionyl-tRNA synthetase  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 16-Jul-1999  
C;Accession: I64113  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630  
A;Accession: I64113  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-682 <TIGR>  
A;Cross-references: GB:U32807; GB:L42023; NID:g1574725; PIDN:AAC22924.1; PID:g1574731  
C;Genetics:  
A;Gene: metG  
C;Superfamily: methionine--tRNA ligase  
C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 44.3%; Score 39; DB 2; Length 682;  
Best Local Similarity 63.6%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLG 12  
|||  
Db 230 WDISRDAPYFG 240

Search completed: June 28, 2001, 11:57:01  
Job time: 242 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:25 ; Search time 22.61 seconds  
(without alignments)  
24.241 Million cell updates/sec

Title: US-09-439-313-562  
Perfect score: 88  
Sequence: 1 DWDTSALAPYLGTQEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	52.3	215	1 YSCR_SALTY	P74890 salmonella
2	43	48.9	206	1 SODM_MYCAV	P47201 mycobacteri
3	43	48.9	626	1 YT17_MYCTU	Q10966 mycobacteri
4	42	47.7	206	1 SODM_MYCLP	O86165 mycobacteri
5	42	47.7	207	1 SODF_MYCTU	P17670 mycobacteri
6	42	47.7	643	1 CIK2_DROME	P08511 drosophila
7	42	47.7	656	1 CIK4_DROME	P08513 drosophila
8	42	47.7	4485	1 DYHG_CHLRE	Q39575 chlamydomon
9	41	46.6	233	1 SODM_YEAST	P00447 saccharomyc
10	41	46.6	415	1 PROA_BACSU	P39821 bacillus su
11	40	45.5	206	1 SODM_MYCLE	P13367 mycobacteri
12	40	45.5	440	1 C21I_HUMAN	Q9nvd3 homo sapien
13	40	45.5	484	1 SGAT_ECOLI	P39301 escherichia
14	40	45.5	492	1 PDII_SCHPO	Q10057 schizosacch
15	39	44.3	320	1 Y054_MYCPN	P75049 mycoplasma
16	39	44.3	453	1 HCAE_ECOLI	Q47139 escherichia
17	39	44.3	453	1 YJ54_ARCFU	O28325 archaeoglob
18	39	44.3	471	1 DAX1_PIG	P79386 sus scrofa
19	39	44.3	676	1 SYM_ECOLI	P00959 escherichia
20	39	44.3	682	1 SYM_HAEIN	P43828 haemophilus
21	39	44.3	727	1 KDGA_RAT	P51556 rattus norv
22	39	44.3	730	1 KDGA_MOUSE	O88673 mus musculu
23	39	44.3	734	1 KDGA_PIG	P20192 sus scrofa
24	39	44.3	735	1 KDGA_HUMAN	P23743 homo sapien
25	38	43.2	562	1 TPA_HUMAN	P00750 homo sapien
26	38	43.2	609	1 UGST_ORYGL	Q42968 oryza glabe
27	38	43.2	609	1 UGST_ORYSA	P19395 oryza sativ
28	38	43.2	767	1 AMYH_SACDI	P04065 saccharomyc
29	38	43.2	912	1 PGCB_BOVIN	Q28062 bos taurus
30	38	43.2	2292	1 POLG_EMCVB	P17593 encephalomy
31	38	43.2	2292	1 POLG_EMCVD	P17594 encephalomy
32	37.5	42.6	260	1 MTX2_CAEEL	P34599 caenorhabdi
33	37	42.0	118	1 YNIF_AZOBR	P25316 azospirillu

34	37	42.0	199	1 NHAA_PSECL	P27764 pseudomonas
35	37	42.0	206	1 SODM_NOCAS	P53651 nocardia as
36	37	42.0	214	1 DGK1_LACAC	Q59483 lactobacill
37	37	42.0	283	1 YP97_CAEEL	Q09245 caenorhabdi
38	37	42.0	295	1 THTR_RAT	P24329 rattus norv
39	37	42.0	296	1 THTR_BOVIN	P00586 bos taurus
40	37	42.0	296	1 THTR_MOUSE	P52196 mus musculu
41	37	42.0	302	1 VG03_HSVII	Q00115 ictaluriid h
42	37	42.0	319	1 HPRK_TREPA	O83600 treponema p
43	37	42.0	358	1 PKX1_HUMAN	P51817 homo sapien
44	37	42.0	379	1 METX_NEIMB	Q9jzq5 neisseria m
45	37	42.0	549	1 AMYG_YEAST	P08019 saccharomyc
46	37	42.0	842	1 LPFC_SALTY	P43662 salmonella
47	37	42.0	993	1 VIA_CMVFN	P17769 cucumber mo
48	37	42.0	993	1 VIA_CMVII	Q83270 cucumber mo
49	37	42.0	993	1 VIA_CMVIX	Q66121 cucumber mo
50	37	42.0	993	1 VIA_CMVNT	O40976 cucumber mo
51	37	42.0	993	1 VIA_CMVO	P20122 cucumber mo
52	37	42.0	993	1 VIA_CMVY	Q83264 cucumber mo
53	37	42.0	1131	1 PMAI_DUNBI	P54211 dunaliella
54	37	42.0	1290	1 BXC1_CLOBO	P18640 clostridium
55	37	42.0	1447	1 YAO6_CAEEL	Q20762 caenorhabdi
56	37	42.0	1459	1 YF1M_CAEEL	Q21874 caenorhabdi
57	37	42.0	2331	1 RRPL_MABVM	P31352 marburg vir
58	37	42.0	2331	1 RRPL_MABVP	P35262 marburg vir
59	36.5	41.5	271	1 YAD8_SCHPO	Q09834 schizosacch
60	36.5	41.5	522	1 LEU1_ECOLI	P09151 escherichia
61	36	40.9	33	1 MYMY_MYTED	P81614 mytilus edu
62	36	40.9	153	1 VG59_BPMD2	O64250 mycobacteri
63	36	40.9	217	1 GTT1_MANSE	P46430 manduca sex
64	36	40.9	256	1 ERG2_NEUCR	Q92254 neurospora
65	36	40.9	262	1 LEC_PHALU	P16300 phaseolus l

ALIGNMENTS

RESULT 1	YSKR_SALTY	STANDARD;	PRT;	215 AA.
ID	YSKR_SALTY	STANDARD;	PRT;	215 AA.
AC	P74890;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	VIRULENCE PROTEIN YSCR.			
GN	YSKR.			
OS	Salmonella typhimurium.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Salmonella.			
OX	NCBI_TaxID=602;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LT2;			
RA	Hensel M., Shea J.E., Baumlner A.J., Gleeson C., Blattner F.R.,			
RA	Holden D.W.;			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-!- SIMILARITY: BELONGS TO THE FLIP/MOPC/SPAP FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; X99944; CAA68199.1; -.			
DR	InterPro; IPR002039; -.			
DR	Pfam; PF00813; FLIP; 1.			
DR	PRINTS; PR00951; FLGBIOSNFLIP.			
DR	PROSITE; PS01060; FLIP_1; 1.			
DR	PROSITE; PS01061; FLIP_2; 1.			

```

KW Virulence; Transmembrane.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
SQ SEQUENCE 215 AA; 24094 MW; 95DDDF04BE275A396 CRC64;

Query Match 52.3%; Score 46; DB 1; Length 215;
Best Local Similarity 70.0%; Pred. No. 1.3;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DWDTSALAPY 10
Db 92 EWDKALAPY 101

RESULT 2
SODM_MYCAV STANDARD; PRT; 206 AA.
AC P47201; 1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
GN SODA OR SOD.
OS Mycobacterium avium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TMC 724;
RX MEDLINE=96276149; PubMed=8692009;
RA Escuyer V.E., Haddad N., Frehel C., Berche P.;
RT "Molecular characterization of a surface-exposed superoxide dismutase
of Mycobacterium avium";
RL Microb. Pathog. 20:41-55(1996).
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U11550; AAB08770.1;
CC HSSP; P17670; LIDS.
CC InterPro; IPR001189;
CC Pfam; PF00081; scdfe; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Manganese.
KW INIT_MET 0 BY SIMILARITY.
FT METAL 27 MANGANESE (BY SIMILARITY).
FT METAL 75 MANGANESE (BY SIMILARITY).
FT METAL 159 MANGANESE (BY SIMILARITY).
FT METAL 163 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 206 AA; 22912 MW; D6B49083AEBF3C98 CRC64;

Query Match 48.9%; Score 43; DB 1; Length 206;
Best Local Similarity 50.0%; Pred. No. 4.1;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTO 14
Db 9 DWDYAAALEPHISGQ 22

```

```

RESULT 3
YTI17_MYCTU STANDARD; PRT; 626 AA.
ID YTI17_MYCTU
AC Q10966; Q10967;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 68.4 KDA PROTEIN RV2917.
GN RV2917 OR MTCY338.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RT Nature 393:537-544(1998).
RL Nature 393:537-544(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z74697; CAA98979.1;
CC Tuberculist; RV2917;
KW Hypothetical protein.
SQ SEQUENCE 626 AA; 68365 MW; C8953B2E52B0B841 CRC64;

Query Match 48.9%; Score 43; DB 1; Length 626;
Best Local Similarity 61.5%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTO 14
Db 68 WDRALVQYLGTO 80

RESULT 4
SODM_MYCLP STANDARD; PRT; 206 AA.
ID SODM_MYCLP
AC O86165;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
GN SODA OR SOD.
OS Mycobacterium lepraemurium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=64667;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HAWAIIAN;
RA Nakamura M.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE

```

CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).  
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D13288; BAA28850.1; -.  
CC HSSP; P17670; IIDS.  
CC InterPro; IPR001189; -.  
CC Pfam; PF00081; sodfe; 1.  
CC PROSITE; PS00088; SOD\_MN; 1.  
KW Oxidoreductase; Manganese.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT METAL 27 27 MANGANESE (BY SIMILARITY).  
FT METAL 75 75 MANGANESE (BY SIMILARITY).  
FT METAL 159 159 MANGANESE (BY SIMILARITY).  
FT METAL 163 163 MANGANESE (BY SIMILARITY).  
SQ SEQUENCE 206 AA; 23031 MW; 92A063471FA9F22F CRC64;  
  
Query Match 47.7%; Score 42; DB 1; Length 206;  
Best Local Similarity 50.0%; Pred. No. 6;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 DWDTSAAPYLGTQ 14  
Db 9 DWDYEALPHISGQ 22  
  
RESULT 5  
SODF\_MYCTU  
ID SODF\_MYCTU STANDARD; PRT; 207 AA.  
AC P17670; P96231;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SUPEROXIDE DISMUTASE [FE] (EC 1.15.1.1).  
GN SODB OR SODA OR SOD OR RV3846 OR MTCY01A6.22C.  
OS Mycobacterium tuberculosis, and Mycobacterium bovis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773, 1765;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=ERDMANN;  
RX MEDLINE=91251768; PubMed=1904126;  
RA Zhang Y.;  
RT "Genetic analysis of superoxide dismutase, the 23 kilodalton antigen  
RT of Mycobacterium tuberculosis.";  
RL Mol. Microbiol. 5:381-391(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [3]

RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=ERDMANN;  
RX MEDLINE=99134360; PubMed=9933629;  
RA Harth G., Horwitz M.A.;  
RT "Export of recombinant Mycobacterium tuberculosis superoxide dismutase  
RT is dependent upon both information in the protein and mycobacterial  
RT export machinery. A model for studying export of leaderless proteins  
RT by pathogenic mycobacteria.";  
RL J. Biol. Chem. 274:4281-4292(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.bovis; STRAIN=BCG / PASTEUR;  
RA Kimble E., Sanderson R.J., Gill R.E.;  
RT "Superoxide dismutase of M. bovis BCG.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBDJ databases.  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
RC SPECIES=M.tuberculosis;  
RX MEDLINE=95182461; PubMed=7877174;  
RA Cooper J.B., McIntyre K., Badasso M.O.; Wood S.P., Zhang Y.,  
RA Garbe T.R., Young D.;  
RT "X-ray structure analysis of the iron-dependent superoxide dismutase  
RT from Mycobacterium tuberculosis at 2.0-A resolution reveals novel  
RT dimer-dimer interactions.";  
RL J. Mol. Biol. 246:531-544(1995).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF MUTANT ALA-152.  
RC SPECIES=M.tuberculosis;  
RX MEDLINE=96244503; PubMed=8674528;  
RA Cooper J.B., Seward S., Erskine P.T., Badasso M.O., Wood S.P.,  
RA Zhang Y., Young D.;  
RT "X-ray structure analysis of an engineered Fe-superoxide dismutase  
RT Gly-Ala mutant with significantly reduced stability to denaturant.";  
RL FEBS Lett. 387:105-108(1996).  
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).  
CC -!- SUBUNIT: HOMOTETRAMER.  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
CC FAMILY.  
CC -!- CAUTION: ALTHOUGH FOUND EXTRACELLULARLY, NO SIGNAL SEQUENCE IS  
CC PRESENT. AN ALTERNATIVE SECRETORY PATHWAY MAY BE USED.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X52861; CAA37042.1; -.  
CC EMBL; Z83864; CAB06220.1; -.  
CC EMBL; AF061030; AAD15824.1; -.  
CC EMBL; AF077406; AAC27527.1; -.  
CC PIR; S10908; S10908.  
CC PIR; S15205; S15205.  
CC PDB; IIDS; 20-DEC-94.  
CC TuberculList; Rv3846; -.  
CC InterPro; IPR001189; -.  
CC Pfam; PF00081; sodfe; 1.  
CC PROSITE; PS00088; SOD\_MN; 1.  
KW Oxidoreductase; Iron; 3D-structure.  
FT METAL 28 28 IRON.  
FT METAL 76 76 IRON.  
FT METAL 160 160 IRON.  
FT METAL 164 164 IRON.  
SQ SEQUENCE 207 AA; 23034 MW; DEE8F5921DABE54A CRC64;  
  
Query Match 47.7%; Score 42; DB 1; Length 207;  
Best Local Similarity 50.0%; Pred. No. 6;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQ 14  
Db 10 DWDYGALEPHISGQ 23

RESULT 6  
CIK2\_DROME STANDARD; PRT; 643 AA.  
ID CIK2\_DROME Q24521;  
AC P08511; Q24521;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN, LATE POPULATION (SHAKER-BETA).  
DE SH.  
GN Drosophila melanogaster (Fruit fly).  
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CANTON-S;  
RX MEDLINE=88296413; PubMed=2456921;  
RA Pongs O., Kecskenethy N., Mueller R., Krah-Jentgens I., Ferrus A.;  
RA Baumann A., Kiltz H.H., Canal I., Llamazares S., Ferrus A.;  
RT "Shaker encodes a family of putative potassium channel proteins in the nervous system of Drosophila."  
RL EMBO J. 7:1087-1096(1988).  
RN [2]  
RP SEQUENCE OF 438-523 FROM N.A.  
RC STRAIN=H4;  
RX MEDLINE=90166523; PubMed=3272175;  
RA Kamb A., Tseng-Crank J., Tanouye M.A.;  
RT "Multiple products of the Drosophila Shaker gene may contribute to potassium channel diversity."  
RL Neuron 1:421-430(1988).  
CC -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A POTASSIUM-SELECTIVE CHANNEL THROUGH WHICH K+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- MISCELLANEOUS: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X07132; CAA30144.1; -;  
CC EMBL; X78908; CAA55519.1; -;  
CC PIR; S00480; S00480.  
CC FlyBase; FBgn003380; Sh.  
CC InterPro; IPR000636; -;  
CC InterPro; IPR003091; -;  
CC Pfam; PF00520; Ion\_trans; 1.  
CC PRINTS; PR00169; KCHANNEL.  
CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW Glycoprotein; Alternative splicing; Multigene family.  
FT TRANSMEM 216 234  
FT TRANSMEM 267 288  
FT TRANSMEM 300 320  
FT TRANSMEM 300 320

FT TRANSMEM 349 367  
FT TRANSMEM 384 403  
FT TRANSMEM 445 466  
FT CARBOHYD 48 90  
FT CARBOHYD 247 251  
FT CARBOHYD 251 251  
FT CARBOHYD 612 612  
FT CONFLICT 501 501  
SQ SEQUENCE 643 AA; 72501 MW; E1F358B555496D2 CRC64;

Query Match 47.7%; Score 42; DB 1; Length 643;  
Best Local Similarity 58.3%; Pred. No. 18;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 SALAPYLGTOEE 16  
Db 537 SAVAPFLGAQQQ 548

RESULT 7  
CIK4\_DROME STANDARD; PRT; 656 AA.  
ID CIK4\_DROME  
AC P08513;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN, LARVAL (SHAKER-EPSILON).  
DE SH.  
GN Drosophila melanogaster (Fruit fly).  
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=88122563; PubMed=2448635;  
RX Schwarzt T.L., Tempel B.L., Papazian D.M., Jan Y.N., Jan L.Y.;  
RA "Multiple potassium-channel components are produced by alternative splicing at the Shaker locus in Drosophila."  
RL Nature 331:137-142(1988).  
RN [2]  
RP SEQUENCE OF 1-349 FROM N.A.  
RC STRAIN=CANTON-S;  
RX MEDLINE=88296413; PubMed=2456921;  
RA Pongs O., Kecskenethy N., Mueller R., Krah-Jentgens I., Baumann A., Kiltz H.H., Canal I., Llamazares S., Ferrus A.;  
RA "Shaker encodes a family of putative potassium channel proteins in the nervous system of Drosophila."  
RL EMBO J. 7:1087-1096(1988).  
RN [3]  
RP FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A POTASSIUM-SELECTIVE CHANNEL THROUGH WHICH K+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- MISCELLANEOUS: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X07134; CAA30146.1; -;





CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) - O(2) + H(2)O(2).  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: X02156; CAA26092.1; -  
CC EMBL: U10400; AAB68939.1; -  
CC EMBL: M24079; AAA35065.1; -  
CC PIR: A00521; DSBYN.  
CC PIR: S46785; S46785.  
CC HSSP: P09214; 1MNG.  
CC SGD: S001050; SOD2.  
CC InterPro: IPR001189; -  
CC Pfam: PF00081; sodfe; 1.  
CC PROSITE: PS00088; SOD\_MN; 1.  
CC Oxidoreductase; Manganese; Mitochondrion; Transit peptide.  
CC TRANSIT 1 26 MITOCHONDRION.  
CC CHAIN 27 233 SUPEROXIDE DISMUTASE [MN].  
CC METAL 52 52 MANGANESE (BY SIMILARITY).  
CC METAL 107 107 MANGANESE (BY SIMILARITY).  
CC METAL 194 194 MANGANESE (BY SIMILARITY).  
CC METAL 198 198 MANGANESE (BY SIMILARITY).  
CC METAL 198 198 MANGANESE (BY SIMILARITY).  
CC SEQUENCE 233 AA; 25774 MW; 88A9391FBB31D06E CRC64;  
CC  
CC Query Match 46.6%; Score 41; DB 1; Length 233;  
CC Best Local Similarity 53.8%; Pred. No. 9.9;  
CC Matches 7; Conservative 1; Mismatches 5; Indels 5; Gaps 0;  
CC  
CC QY 2 WDTSALAPYLGTQ 14  
CC II II II I  
CC DB 35 WDFGALEPYISGQ 47  
CC  
CC RESULT 10  
CC PROA\_BACSU STANDARD; PRT; 415 AA.  
CC ID PROA\_BACSU STANDARD; PRT; 415 AA.  
CC AC P39821; O35032;  
CC DT 01-FEB-1995 (Rel. 31, Created)  
CC DT 30-MAY-2000 (Rel. 39, Last sequence update)  
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
CC DE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR) (EC 1.2.1.41) (GLUTAMATE-5-  
CC SEMIALDEHYDE DEHYDROGENASE) (GLUTAMYL-GAMMA-SEMIALDEHYDE  
CC DEHYDROGENASE).  
CC GN PROA.  
CC OS Bacillus subtilis.  
CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
CC OC Bacillus/Staphylococcus group; Bacillus.  
CC OX NCBI\_TaxID=1423;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN=168;  
CC RX MEDLINE=94364946; PubMed=8083159;  
CC RA Ogura M., Kawata-Mukai M., Itaya M., Takio K., Tanaka T.;  
CC RT "Multiple copies of the prob gene enhance degS-dependent  
CC extracellular protease production in Bacillus subtilis."  
CC RL J. Bacteriol. 176:5673-5680(1994).  
CC RN [2]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN=168;  
CC RA Devine K.M.;  
CC RT "Sequence of the Bacillus subtilis genome between xlyA and ykor.";  
CC RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE 5-SEMIALDEHYDE + ORTHOPHOSPHATE +

CC NADP(+) - L-GAMMA-GLUTAMYL 5-PHOSPHATE + NADPH.  
CC -1- PATHWAY: SECOND STEP IN PROLINE BIOSYNTHESIS PATHWAY.  
CC -1- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE  
CC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: D26044; BAA05045.1; -  
CC EMBL: AJ002571; CAA05592.1; -  
CC EMBL: Z99110; CAB13170.1; -  
CC Subtilist; BG10964; PROA.  
CC InterPro: IPR000965; -  
CC InterPro: IPR002086; -  
CC Pfam: PF00171; aldedh; 1.  
CC PROSITE: PS01223; PROA; 1.  
CC Oxidoreductase; Proline biosynthesis; NADP.  
CC CONFLICT 108 108 E -> Q (IN REF. 1).  
CC CONFLICT 174 174 A -> T (IN REF. 1).  
CC CONFLICT 271 271 H -> N (IN REF. 1).  
CC CONFLICT 359 359 R -> A (IN REF. 1).  
CC CONFLICT 374 415 FEFGAEIGISTOKLHARGPMGLPALSTKTYIKGTGQIR  
CC E -> FGIRLSRNRHQHAKASCKRTDGA SCTDFYKIHH  
CC (IN REF. 1).  
CC SEQUENCE 415 AA; 45336 MW; 8CA4B0D35F9F62D0 CRC64;  
CC  
CC Query Match 46.6%; Score 41; DB 1; Length 415;  
CC Best Local Similarity 50.0%; Pred. No. 17;  
CC Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
CC  
CC QY 1 DWDTSLAPYLGTQ 14  
CC II II II I  
CC DB 306 DWETEFLAPVLSVK 319  
CC  
CC RESULT 11  
CC SODM\_MYCLE STANDARD; PRT; 206 AA.  
CC AC P13367;  
CC DT 01-JAN-1990 (Rel. 13, Created)  
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
CC DT 01-OCT-1996 (Rel. 34, Last annotation update)  
CC DE SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).  
CC GN SODA.  
CC OS Mycobacterium leprae.  
CC OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
CC OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
CC OX NCBI\_TaxID=1769;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC MEDLINE=90045970; PubMed=2682526;  
CC RA Thangaraj H.S., Lamb F.I., Davis E.O., Colston M.J.;  
CC RT "Nucleotide and deduced amino acid sequence of Mycobacterium leprae  
CC manganese superoxide dismutase".  
CC RL Nucleic Acids Res. 17:8378-8378(1989).  
CC RN [2]  
CC RP SEQUENCE FROM N.A.  
CC RC MEDLINE=90256282; PubMed=1692812;  
CC RA Thangaraj H.S., Lamb F.I., Davis E.O., Jenner P.J., Jeyakumar L.H.,  
CC Colston M.J.;  
CC RT "Identification, sequencing, and expression of Mycobacterium leprae  
CC superoxide dismutase, a major antigen."  
CC RL Infect. Immun. 58:1937-1942(1990).  
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) - O(2) + H(2)O(2).  
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE

CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X16453; CAA34472.1; -  
DR PIR; S06599; S06599.  
DR PIR; A37212; A37212.  
DR HSSP; PI7670; 1IDS.  
DR InterPro; IPR001189; -  
DR pfam; PF00081; sodfe; 1.  
DR PROSITE; PS00088; SOD\_MN; 1.  
KW Oxidoreductase; Manganese.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT METAL 27 27 MANGANESE (BY SIMILARITY).  
FT METAL 75 75 MANGANESE (BY SIMILARITY).  
FT METAL 159 159 MANGANESE (BY SIMILARITY).  
FT METAL 163 163 MANGANESE (BY SIMILARITY).  
SQ SEQUENCE 206 AA; 23027 MW; B02B96EB433F6FF0 CRC64;

Query Match 45.5%; Score 40; DB 1; Length 206;  
Best Local Similarity 54.5%; Pred. No. 13;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11  
||| :||| :  
Db 9 DWDYAALPHI 19

RESULT 12  
C21I\_HUMAN STANDARD; PRT; 440 AA.  
AC Q9NVD3;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PUTATIVE PROTEIN C21ORF18.  
GN C21ORF18.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO S.POMBE C1223.04C.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AK001660; BAA91819.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 440 AA; 50416 MW; 9EBCAA05397BE287 CRC64;

Query Match 45.5%; Score 40; DB 1; Length 440;  
Best Local Similarity 88.9%; Pred. No. 27;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 DTSALAPYL 11  
||| :||| :  
Db 224 DTCALAPYL 232

RESULT 13  
SGAT\_ECOLI STANDARD; PRT; 484 AA.  
AC P39301;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PUTATIVE TRANSPORT PROTEIN SGAT.  
GN SGAT.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=95334362; PubMed=7610040;  
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
RA Blattner F.R.;  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
RT region from 92.8 through 100 minutes.";  
RL Nucleic Acids Res. 23:2105-2119(1995).  
RN [2]

RP REVISIONS TO C-TERMINUS.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]

RP DISCUSSION OF SEQUENCE.  
RA Reizer J., Charbit A., Reizer A., Saier M.H. Jr.;  
RT "Novel phosphotransferases system genes revealed by bacterial genome  
RT analysis: operons encoding homologues of sugar-specific permease  
RT domains of the phosphotransferase system and pentose catabolic  
RT enzymes.";  
RL Genome Sci. Technol. 1:53-75(1996).  
CC -!- FUNCTION: COULD ACT AS THE TRANSPORT PROTEIN FOR THE UNKNOWN  
CC PENTITOL SUBSTRATE OF THE SGA OPERON.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -!- SIMILARITY: TO M.PNEUMONIAE SGAT HOMOLOG.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U14003; AAA97089.1; -  
DR EMBL; AE000491; AAC77150.1; -  
DR EcoGene; EG12493; sgat.  
KW Transport; Transmembrane; Inner membrane.  
FT TRANSMEM 12 32 POTENTIAL.  
FT TRANSMEM 33 53 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 120 140 POTENTIAL.  
FT TRANSMEM 160 180 POTENTIAL.  
FT TRANSMEM 252 272 POTENTIAL.  
FT TRANSMEM 282 302 POTENTIAL.  
FT TRANSMEM 335 355 POTENTIAL.

FT TRANSMEM 357 377 POTENTIAL.  
FT TRANSMEM 398 418 POTENTIAL.  
FT TRANSMEM 420 440 POTENTIAL.  
FT TRANSMEM 446 466 POTENTIAL.  
FT CONFLICT 470 484 RAEDAERKQAEQSA -> AOKKMKNNWQNSLLNKEF  
SQ SEQUENCE 484 AA; 52950 MW; 2CBEBD0044BC6CAC CRC64;  
  
Query Match 45.5%; Score 40; DB 1; Length 484;  
Best Local Similarity 46.7%; Pred. No. 30; Indels 0; Gaps 0;  
Matches 7; Conservative 3; Mismatches 5; Indels 5; Gaps 0;  
  
QY 2 WDTSAALAPYLGTQEE 16  
Db 220 WIAYKVAPFLGKKEE 234  
  
RESULT 14  
PDII1\_SCHPO STANDARD; PRT: 492 AA.  
AC Q10057;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PUTATIVE PROTEIN DISULFIDE ISOMERASE C1F5.02 PRECURSOR (EC 5.3.4.1).  
GN SPAC1F5.02.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetes;  
OC Schizosaccharomycetes;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Gentles S., Odell C., Churcher C.M., Barrell B.G., Rajandream M.A.,  
RA Walsh S.V.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING  
CC DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROLYL  
CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN  
CC DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.  
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z68136; CAA92230.1; -  
CC HSSP; P07237; 1MEK.  
CC InterPro; IPR000063; -  
CC InterPro; IPR000886; -  
CC Pfam; PF00085; thioered; 2.  
CC PRINTS; PR00421; THIOREDOXIN.  
CC PROSITE; PS00014; ER\_TARGET; 1.  
CC PROSITE; PS00194; THIOREDOXIN; 2.  
KW Hypothetical protein; Redox-active center; Isomerase;  
KW Endoplasmic reticulum; Repeat; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 492  
FT  
FT PUTATIVE PROTEIN DISULFIDE ISOMERASE  
FT C1F5.02.  
FT  
FT REDOX-ACTIVE (BY SIMILARITY).  
FT DISULFID 51 54  
FT DISULFID 385 388 REDOX-ACTIVE (BY SIMILARITY).  
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SITE 489 492 PREVENT SECRETION FROM ER (POTENTIAL).  
SQ SEQUENCE 492 AA; 54880 MW; C50B592365666667 CRC64;

Query Match 45.5%; Score 40; DB 1; Length 492;  
Best Local Similarity 41.7%; Pred. No. 30; Indels 0; Gaps 0;  
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 DWDTSALAPYLIG 12  
Db 215 DWDPAIADFIG 226  
  
RESULT 15  
Y054\_MYCPN STANDARD; PRT: 320 AA.  
AC P75049;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN MG054 HOMOLOG (D09\_ORF320).  
GN MPN067 OR MP087.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129; PubMed=8948633;  
RX MEDLINE=97105885; Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RA Himmelfreid R., Herrmann R.;  
RA "Complete sequence analysis of the genome of the bacterium Mycoplasma  
RT pneumoniae"; Res. 24:4420-4449(1996).  
RL Nucleic Acids; Res. 24:4420-4449(1996).  
CC -1- SIMILARITY: BELONGS TO THE NUSG FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE000011; AAB95735.1; -  
CC Hypothetical protein.  
KW SEQUENCE 320 AA; 36107 MW; ECE3C626C8EAE63F CRC64;  
  
Query Match 44.3%; Score 39; DB 1; Length 320;  
Best Local Similarity 61.5%; Pred. No. 29; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 3 DTSALAPYLGTQE 15  
Db 234 DPSVLIPYLKYE 246  
  
RESULT 16  
HCAE\_ECOLI STANDARD; PRT: 453 AA.  
AC Q47139; P77590;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE 3-PHENYLPROPIONATE DIOXYGENASE ALPHA SUBUNIT (EC 1.14.1.-)  
GN (DIOXYGENIN ALPHA SUBUNIT).  
GN HCAE OR PHDCl OR HCAA OR HCAAL OR DIGA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.



RC STRAIN=K12;  
RA Turlin E., Gasser F., Biville F.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97349980; PubMed=9205837;  
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features.";  
RL DNA Res. 4:91-113(1997).  
CC -!- FUNCTION: PART OF MULTICOMPONENT 3-PHENYLPROPIONATE DIOXYGENASE.  
CC CONVERTS 3-PHENYLPROPIONIC ACID (PP) INTO CIS-3-(3-CARBOXYETHYL)-  
CC 3,5-CYCLOHEXADIENE-1,2-DIOL (PP-DIHYDRODIOL).  
CC -!- COFACTOR: PROBABLY BINDS A 2FE-2S GROUP AND AN IRON ATOM.  
CC -!- PATHWAY: CATABOLISM OF 3-PHENYLPROPIONIC ACID.  
CC -!- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE  
CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (HCAE AND HCAF), A  
CC FERREDOXIN (HCAC) AND A FERREDOXIN REDUCTASE (HCAD).  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL RING-HYDROXYLATING  
CC DIOXYGENASE ALPHA SUBUNIT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z37966; CAA86018.1; -.  
DR EMBL; AE000340; AAC75591.1; -.  
DR EMBL; D90884; BAA16441.1; -.  
DR EcoGene; EGI3456; hcae.  
DR InterPro; IPR001281; -.  
DR InterPro; IPR001663; -.  
DR Pfam; PF00355; Rieske; 1.  
DR Pfam; PF00848; Ring\_hydroxyl\_A; 1.  
DR PRINTS; PR00090; RINGDIOXGNASE.  
DR PROSITE; PS00570; RING\_HYDROXYL\_ALPHA; 1.  
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Iron-sulfur; Iron;  
KW Dioxygenase; NAD.  
FT METAL 85 85 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 87 87 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 105 105 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 108 108 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 213 213 IRON (BY SIMILARITY).  
FT METAL 218 218 IRON (BY SIMILARITY).  
FT CONFLICT 20 20 V -> A (IN REF. 1).  
FT CONFLICT 384 453 GHRARNKLCLEMGLEQEKRRDDGIPGITYNIFSETAARGM  
FT YQRWADLLSSSEWQEVLDKTAAYQQEVMK -> ATAPATAN  
FT CVWKGLVRKSAATFAFLALLTISFQKPLVLECTNAGPIF  
FT (IN REF. 1).  
SQ SEQUENCE 453 AA; 51109 MW; 02535BF5F47643FD CRC64;

Query Match 44.3%; Score 39; DB 1; Length 453;

Best Local Similarity 64.3%; Pred. No. 41;  
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;  
QY 1 DWDTSA--LAPYLIG 12  
:||||| | |||  
Db 155 NWDTSAAPGLRDYLG 168  
RESULT 17  
YJ54\_ARCFU STANDARD; PRT; 453 AA.  
ID YJ54\_ARCFU  
AC O28325;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE AMIDASE AF1954 (EC 3.5.1.4).  
GN AF1954.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
OC Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -!- CATALYTIC ACTIVITY: A MONOCARBOXYLIC ACID AMIDE + H(2)O -  
CC A MONOCARBOXYLATE + NH(3).  
CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY. STRONG, TO GATA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000968; AAB89301.1; -.  
DR TIGR; AF1954; -.  
DR InterPro; IPR000120; -.  
DR Pfam; PF01425; Amidase; 1.  
DR PROSITE; PS00571; AMIDASES; 1.  
KW Hypothetical protein; Hydrolase.  
SQ SEQUENCE 453 AA; 50062 MW; 72E8D6A2A535F483 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 453;  
Best Local Similarity 50.0%; Pred. No. 41;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLIG 12  
:|: | |||  
Db 313 EWEKVAFPYPYLIG 324

RESULT 18  
DAXL\_PIG STANDARD; PRT; 471 AA.  
ID DAXL\_PIG  
AC P79386;

DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ORPHAN NUCLEAR RECEPTOR\_DAX-1.  
GN NROB1 OR AUCH OR DAX1.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal gland;  
RX MEDLINE=98113985; PubMed=9453240;  
RA Parma P., Pailhoux E., Puissant C., Cotinot C.;  
RT "Porcine Dax-1 gene: isolation and expression during gonadal  
RL development".  
RN Mol. Cell. Endocrinol. 135:49-58(1997).  
RN [2]  
RP SEQUENCE OF 1-390 FROM N.A.  
RA Behdjani R., Silversides D.W.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: RECEPTOR THAT MAY BE A COMPONENT OF A CASCADE REQUIRED  
CC FOR DEVELOPMENT OF STEROIDGENIC TISSUES. ACTS AS A DOMINANT  
CC NEGATIVE REGULATOR OF TRANSCRIPTION MEDIATED BY THE RETINOIC ACID  
CC RECEPTOR (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
CC NRO SUBFAMILY. LACKS DNA-BINDING REGION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U82466; AAB81101.1; --  
DR EMBL; AF019044; AAB70254.1; --  
DR InterPro; IPR000536; --  
DR Pfam; PF00104; hormone\_rec; 1.  
KW Receptor; Nuclear protein; Transcription regulation; Repressor;  
KW Repeat.  
FT DOMAIN 1 253 4 X 67 AA TANDEM REPEATS.  
FT REPEAT 1 67  
FT REPEAT 68 134  
FT REPEAT 135 201  
FT REPEAT 202 253  
FT DOMAIN 254 471 LIGAND-BINDING (BY SIMILARITY).  
FT CONFLICT 124 125 GR -> AG (IN REF. 2).  
FT CONFLICT 243 244 AQ -> OR (IN REF. 2).  
FT CONFLICT 266 266 L -> V (IN REF. 2).  
FT CONFLICT 289 289 S -> T (IN REF. 2).  
SQ SEQUENCE 471 AA; 52128 MW; 8EE7D133A4677950 CRC64;  
  
Query Match 44.3%; Score 39; DB 1; Length 471;  
Best Local Similarity 53.3%; Pred. No. 42;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
  
OY 2 WDTSALAPYLGTQEE 16  
DB 173 WDRSYCAQRLGAREE 187  
  
RESULT 19  
SYM\_ECOLI  
ID SYM\_ECOLI STANDARD; PRT; 676 AA.  
AC P00959;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)

DE (METRS).  
GN METG.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=85054627; PubMed=6094501;  
RX Dardel F., Fayat G., Blanquet S.;  
RA "Molecular cloning and primary structure of the Escherichia coli  
RT methionyl-tRNA synthetase gene."  
RN J. Bacteriol. 160:1115-1122(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=91080852; PubMed=2259334;  
RA Dardel F., Panvert M., Fayat G.;  
RT "Transcription and regulation of expression of the Escherichia coli  
RL methionyl-tRNA synthetase gene."  
RN Mol. Gen. Genet. 223:121-133(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / BHB2600;  
RX Church G.M.;  
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,  
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MGL655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RN Science 277:1453-1474(1997).  
RN [5]  
RP SEQUENCE OF 1-565 FROM N.A. AND PARTIAL SEQUENCE.  
RX MEDLINE=83079258; PubMed=6756915;  
RA Barker D.G., Ebel J.-P., Jakes R., Bruton C.J.;  
RT "Methionyl-tRNA synthetase from Escherichia coli. Primary structure  
RL of the active crystallised tryptic fragment."  
RN Eur. J. Biochem. 127:449-457(1982).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=82192427; PubMed=7042987;  
RA Zelwer C., Risler J.-L., Brunie S.;  
RT "Crystal structure of Escherichia coli methionyl-tRNA synthetase at  
RN 2.5-A resolution."  
RN J. Mol. Biol. 155:63-81(1982).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=91073404; PubMed=2254937;  
RA Brunie S., Zelwer C., Risler J.-L.;  
RT "Crystallographic study at 2.5-A resolution of the interaction of  
RN methionyl-tRNA synthetase from Escherichia coli with ATP."  
RN J. Mol. Biol. 216:411-424(1990).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF 1-552.  
RX MEDLINE=20069948; PubMed=10600385;  
RA Mechulam Y., Schmitt E., Maveyraud L., Zelwer C., Nureki O.,  
RA Yokoyama S., Konno M., Blanquet S.;  
RT "Crystal structure of Escherichia coli methionyl-tRNA synthetase  
RN highlights species-specific features."  
RN J. Mol. Biol. 294:1287-1297(1999).  
RN [9]  
RP STRUCTURE BY NMR OF 138-163.  
RX MEDLINE=93294859; PubMed=8515466;  
RA Fourmy D., Dardel F., Blanquet S.;  
RT "Methionyl-tRNA synthetase zinc binding domain. Three-dimensional  
RN structure and homology with rubredoxin and gag retroviral proteins."  
RN J. Mol. Biol. 231:1078-1089(1993).

RN RP
 RX MEDLINE=91084494; PubMed=1702021;
 RA Hountondji C., Schmitter J.-M., Beauvallet C., Blanquet S.;
 RT "Mapping of the active site of Escherichia coli methionyl-tRNA
 synthetase: identification of amino acid residues labeled by
 RT periodate-oxidized tRNA(fMet) molecules having modified lengths at
 RT the 3'-acceptor end.";
 RL Biochemistry 29:8190-8198(1990).
 RN [11]
 RP MUTAGENESIS.
 RX MEDLINE=92070503; PubMed=1959615;
 RA Fourmy D., Mechulam Y., Brunie S., Blanquet S., Fayat G.;
 RT "Identification of residues involved in the binding of methionine by
 RT Escherichia coli methionyl-tRNA synthetase.";
 RL FEBS Lett. 292:259-263(1991).
 RN [12]
 RP MUTAGENESIS OF ZINC-LINGANDS.
 RX MEDLINE=93294858; PubMed=8515465;
 RA Fourmy D., Meinnel T., Mechulam Y., Blanquet S.;
 RT "Mapping of the zinc binding domain of Escherichia coli
 RT methionyl-tRNA synthetase.";
 RL J. Mol. Biol. 231:1068-1077(1993).
 RN [13]
 RP REVIEW.
 RX MEDLINE=91129305; PubMed=2126467;
 RA Meinnel T., Mechulam Y., Dardel F., Schmitter J.M., Hountondji C.,
 RA Brunie S., Dessen P., Fayat G., Blanquet S.;
 RT "Methionyl-tRNA synthetase from E. coli -- a review.";
 RL Biochimie 72:625-632(1990).
 CC -!- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
 CC TRNA(FMET) AMINOACYLATION.
 CC -!- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) -> AMP +
 CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -!- COFACTOR: BINDS ONE ZINC ION.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -!- STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to [licensed@isb-sib.ch](mailto:licensed@isb-sib.ch)).
 CC -----
 CC EMBL; K02671; AAA24161.1; -.
 CC DR EMBL; X55791; CAA39315.1; -.
 CC DR EMBL; U00007; AAA60526.1; ALT\_INIT.
 CC DR EMBL; AE000300; AAC75175.1; -.
 CC DR PIR; S14427; SYECMT.
 CC DR PIR; A35821; A35821.
 CC DR PIR; S11949; S11949.
 CC DR PDB; 1MEA; 31-JAN-94.
 CC DR PDB; 1QOT; 01-JAN-00.
 CC DR SWISS-2DPAGE; P00959; COLI.
 CC DR ECO2DBASE; F072.0; 6TH EDITION.
 CC DR ECO2DBASE; G072.0; 6TH EDITION.
 CC DR EcoGene; EG10586; metG.
 CC DR InterPro; IPR001412; -.
 CC DR InterPro; IPR002300; -.
 CC DR InterPro; IPR002304; -.
 CC DR InterPro; IPR002547; -.
 CC DR Pfam; PF00133; tRNA-synt\_1; 1.
 CC DR Pfam; PF01588; tRNA\_bind; 1.
 CC DR PRINTS; PR01041; TRNASYNTHMET.
 CC DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW trRNA-binding; Metal-binding; Zinc; 3D-structure.
 FT INIT\_MET 0
 FT SITE 14 24 "HIGH" REGION.
 FT SITE 332 336 "KMSKS" REGION.
 FT DOMAIN 580 674 TRNA BINDING.
 FT METAL 145 145 ZINC.
 FT METAL 148 148 ZINC.
 FT METAL 158 158 ZINC.
 FT METAL 161 161 ZINC.
 FT BINDING 335 335 ATP.
 FT MUTAGEN 335 335 K->Q,A,E,R: LOSS OF ACTIVITY.
 FT STRAND 141 141
 FT TURN 142 143
 FT STRAND 144 144
 FT TURN 146 148
 SQ SEQUENCE 676 AA; 76123 MW; 917CD98FE70E7728 CRC64;

 Query Match 44.3%; Score 39; DB 1; Length 676;
 Best Local Similarity 63.6%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

 QY 2 WDTSAALPYLG 12
 || | ||| |
 Db 229 WDISRDAPYFG 239

 RESULT 20
 SYM\_HAEIN
 ID SYM\_HAEIN STANDARD; PRT; 682 AA.
 AC P43828;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
 DE (MTRS).
 GN METG OR H11276.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI\_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -!- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
 CC TRNA(FMET) AMINOACYLATION.
 CC -!- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) -> AMP +
 CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -!- STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way



```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32807; AAC22924.1; -
DR HSSP; P00959; IMED.
DR TIGR; H11276; -
DR InterPro; IPR001412; -
DR InterPro; IPR002300; -
DR InterPro; IPR002304; -
DR InterPro; IPR002547; -
DR Pfam; PF00133; tRNA-synt_1; 1.
DR Pfam; PF01588; tRNA_bind; 1.
DR PRINTS; PR01041; TRNASYNTHMET.
DR PROSITE; PS00178; AA-TRNA_LIGASE-I; 1.
KW Aminoacyl-tRNA synthetase; protein biosynthesis; Ligase; ATP-binding;
KW tRNA-binding; Metal-binding; Zinc.
FT SITE 15 25
FT SITE 331 335 "HIGH" REGION.
FT BINDING 334 334 KMSKS" REGION.
FT DOMAIN 586 680 ATP (BY SIMILARITY).
FT METAL 146 146 TRNA BINDING.
FT METAL 149 149 ZINC (BY SIMILARITY).
FT METAL 159 159 ZINC (BY SIMILARITY).
FT METAL 162 162 ZINC (BY SIMILARITY).
SQ SEQUENCE 682 AA; 77023 MW; 61F07EBDFD2A3CE2 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 682;
Best Local Similarity 63.6%; Pred. NO. 61;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYL 12
DB 230 WDISRDAPYFG 240

RESULT 21
ID KDGA_RAT STANDARD; PRT; 727 AA.
AC P51556;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIACYLGLYCEROL KINASE, ALPHA (EC 2.7.1.107) (DIGLYCERIDE KINASE) (DGK-
DE ALPHA) (DAG KINASE ALPHA) (80 KDA DIACYLGLYCEROL KINASE).
GN DGKA OR DAGKI OR DAGK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain; PubMed=1339302;
RX MEDLINE=93095720;
RA Goto K., Watanabe M., Kondo H., Yuasa H., Sakane F., Kanoh H.;
RT "Gene cloning, sequence, expression and in situ localization of 80
RT kDa diacylglycerol kinase specific to oligodendrocyte of rat brain.";
RL Brain Res. Mol. Brain Res. 16:75-87(1992).
CC -!- FUNCTION: UPON CELL STIMULATION CONVERTS THE SECOND MESSENGER
CC DIACYLGLYCEROL INTO PHOSPHATIDATE, INITIATING THE RESYNTHESIS
CC OF PHOSPHATIDYLINOSITOLS AND ATTENUATING PROTEIN KINASE C
CC ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL = ADP +
CC 1,2-DIACYLGLYCEROL 3-PHOSPHATE.
CC -!- ENZYME REGULATED BY CALCIUM AND PHOSPHATIDYL SERINE.
CC -!- PHOSPHORYLATED BY PROTEIN KINASE C (BY SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: LYMPHOCYTES AND OLIGODENDROGLIAL CELLS.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
```

```
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S49760; AAB24434.1; -
DR InterPro; IPR000756; -
DR InterPro; IPR001206; -
DR InterPro; IPR002048; -
DR InterPro; IPR002219; -
DR Pfam; PF00609; DAGKA; 1.
DR Pfam; PF00781; DAGKC; 1.
DR Pfam; PF00130; DAG_PE-bind; 2.
DR Pfam; PF00036; ehand; 2.
DR PRINTS; PR00008; DAGPEDOMAIN.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
DR TRANSFERASE; Kinase; Calcium-binding; Phorbol-ester binding;
KW Multigene family.
FT CA_BIND 121 132 SITE 1 (PROBABLE).
FT CA_BIND 166 177 SITE 2 (PROBABLE).
FT DOMAIN 204 251 PHORBOL-ESTER AND DAG BINDING (BY
FT DOMAIN 268 317 SIMILARITY).
FT DOMAIN 367 492 PHORBOL-ESTER AND DAG BINDING (BY
FT DOMAIN 512 693 SIMILARITY).
FT SEQUENCE 727 AA; 82198 MW; B5A248ADD2F61C1D CRC64;

Query Match 44.3%; Score 39; DB 1; Length 727;
Best Local Similarity 63.6%; Pred. NO. 65;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11
DB 148 DWDVSELRPIL 158

RESULT 22
ID KDGA_MOUSE STANDARD; PRT; 730 AA.
AC O88673;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIACYLGLYCEROL KINASE, ALPHA (EC 2.7.1.107) (DIGLYCERIDE KINASE) (DGK-
DE ALPHA) (DAG KINASE ALPHA) (80 KDA DIACYLGLYCEROL KINASE).
GN DGKA OR DAGK1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC San Juan M.A., Carrera A.C., Merida I.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: UPON CELL STIMULATION CONVERTS THE SECOND MESSENGER
CC DIACYLGLYCEROL INTO PHOSPHATIDATE, INITIATING THE RESYNTHESIS
CC OF PHOSPHATIDYLINOSITOLS AND ATTENUATING PROTEIN KINASE C
CC ACTIVITY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL = ADP +
CC 1,2-DIACYLGLYCEROL 3-PHOSPHATE.
CC -!- ENZYME REGULATED BY CALCIUM AND PHOSPHATIDYL SERINE.
CC -!- PHOSPHORYLATED BY PROTEIN KINASE C (BY SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
```



CC FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
CC BINDING DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF085219; AAC33483.1; -.  
CC MGD; MGI:102952; Dagk1.  
CC InterPro; IPR000756; -.  
CC InterPro; IPR001206; -.  
CC InterPro; IPR002048; -.  
CC InterPro; IPR002219; -.  
CC Pfam; PF00609; DAGKa; 1.  
CC Pfam; PF00781; DAGKC; 1.  
CC Pfam; PF00130; DAG\_PE-bind; 2.  
CC Pfam; PF0036; efhand; 2.  
CC PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
CC PROSITE; PS50081; DAG\_PE\_BIND\_DOM\_2; 2.  
CC PROSITE; PS00018; EF\_HAND; 2.  
CC Transferase; Kinase; Calcium-binding; Phorbol-ester binding;  
KW Multigene family.  
FT CA\_BIND 124 135 SITE 1 (PROBABLE).  
FT CA\_BIND 169 180 SITE 2 (PROBABLE).  
FT DOMAIN 207 254 PHORBOL-ESTER AND DAG BINDING (BY  
FT SIMILARITY).  
FT DOMAIN 271 320 PHORBOL-ESTER AND DAG BINDING (BY  
FT SIMILARITY).  
FT DOMAIN 370 495 CATALYTIC-A (POTENTIAL).  
FT DOMAIN 515 696 CATALYTIC-B (POTENTIAL).  
SQ SEQUENCE 730 AA; 82791 MW; 277D8975961599DE CRC64;

Query Match 44.3%; Score 39; DB 1; Length 730;  
Best Local Similarity 63.6%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11  
||| | | | |  
Db 151 DWDVSELRPIL 161  
  
RESULT 23  
KDGA\_PIG STANDARD; PRT; 734 AA.  
AC P20192;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE DIACYLGLYCEROL KINASE, ALPHA (EC 2.7.1.107) (DIGLYCERIDE KINASE) (DGK-  
DE ALPHA) (DAG KINASE ALPHA) (80 KDA DIACYLGLYCEROL KINASE).  
GN DGKA OR DAGK1 OR DAGK.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Lymphocytes;  
RX MEDLINE=90190867; PubMed=2156169;  
RA Sakane F., Yamada K., Kanoh H., Yokoyama C., Tanabe T.;  
RT \*Porcine diacylglycerol kinase sequence has zinc finger and E-F hand  
RT motifs\*;  
RL Nature 344:345-348(1990).  
RN [2]  
RP CATALYTIC DOMAIN, AND MUTAGENESIS.  
RX MEDLINE=96404909; PubMed=8809050;

RA Sakane F., Kai M., Wada I., Imai S.-I., Kanoh H.;  
RT \*The C-terminal part of diacylglycerol kinase alpha lacking zinc  
RT fingers serves as a catalytic domain\*;  
RL Biochem. J. 318:583-590(1996).  
CC -!- FUNCTION: UPON CELL STIMULATION CONVERTS THE SECOND MESSENGER  
CC DIACYLGLYCEROL INTO PHOSPHATIDATE, INITIATING THE RESYNTHESIS  
CC OF PHOSPHATIDYLINOSITOLS AND ATTENUATING PROTEIN KINASE C  
CC ACTIVITY.  
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL = ADP +  
CC 1,2-DIACYLGLYCEROL 3-PHOSPHATE.  
CC -!- ENZYME REGULATION: STIMULATED BY CALCIUM AND PHOSPHATIDYLSELINE.  
CC PHOSPHORYLATED BY PROTEIN KINASE C.  
CC -!- SUBUNIT: MONOMER.  
CC -!- TISSUE SPECIFICITY: LYMPHOCYTES AND OLIGODENDROGLIAL CELLS.  
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE  
CC FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
CC BINDING DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X53256; CAA37347.1; -.  
CC PIR; S09156; S09156.  
CC InterPro; IPR000756; -.  
CC InterPro; IPR001206; -.  
CC InterPro; IPR002048; -.  
CC InterPro; IPR002219; -.  
CC Pfam; PF00609; DAGKa; 1.  
CC Pfam; PF00781; DAGKC; 1.  
CC Pfam; PF00130; DAG\_PE-bind; 2.  
CC Pfam; PF0036; efhand; 2.  
CC PRINTS; PR00008; DAGPEDOMAIN.  
CC PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
CC PROSITE; PS50081; DAG\_PE\_BIND\_DOM\_2; 2.  
CC PROSITE; PS00018; EF\_HAND; 2.  
KW Transferase; Kinase; Calcium-binding; Phorbol-ester binding;  
KW Multigene family.  
FT CA\_BIND 122 133 SITE 1 (PROBABLE).  
FT CA\_BIND 167 178 SITE 2 (PROBABLE).  
FT DOMAIN 205 252 PHORBOL-ESTER AND DAG BINDING (BY  
FT SIMILARITY).  
FT DOMAIN 269 318 PHORBOL-ESTER AND DAG BINDING (BY  
FT SIMILARITY).  
FT DOMAIN 373 499 CATALYTIC-A (POTENTIAL).  
FT DOMAIN 519 700 CATALYTIC-B (POTENTIAL).  
FT MUTAGEN 248 248 K->R: NO DECREASE IN ACTIVITY.  
FT MUTAGEN 383 383 K->N: NO DECREASE IN ACTIVITY.  
FT MUTAGEN 395 395 K->N: NO DECREASE IN ACTIVITY.  
FT MUTAGEN 483 483 K->N: NO DECREASE IN ACTIVITY.  
FT MUTAGEN 492 492 K->R: NO DECREASE IN ACTIVITY.  
FT MUTAGEN 554 554 K->N: NO DECREASE IN ACTIVITY.  
SQ SEQUENCE 734 AA; 82606 MW; 711C2E66FB4B4E80 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 734;  
Best Local Similarity 63.6%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11  
||| | | | |  
Db 149 DWDVSELRPIL 159  
  
RESULT 24  
KDGA\_HUMAN STANDARD; PRT; 735 AA.  
ID KDGA\_HUMAN

P23743; O75481; O75482; O75483;  
 01-NOV-1991 (Rel. 20, Created)  
 01-NOV-1995 (Rel. 32, Last sequence update)  
 01-OCT-2000 (Rel. 40, Last annotation update)  
 DIACYLGLYCEROL KINASE, ALPHA (EC 2.7.1.107) (DIGLYCERIDE KINASE) (DGK-  
 ALPHA) (DAG KINASE ALPHA) (80 KDA DIACYLGLYCEROL KINASE).  
 DGKA OR DAGK1 OR DAGK.  
 Homo sapiens (Human).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 SEQUENCE FROM N.A.  
 TISSUE=Lymphocytes;  
 MEDLINE=91085550; PubMed=2175712;  
 Schaalp D., de Wit J., van der Wal J., Vandekerckhove J., van  
 Damme J., Gussow D., Ploegh H.L., van Blitterswijk W.J.,  
 van der Bend R.L.;  
 "Purification, cDNA-cloning and expression of human diacylglycerol  
 kinase.";  
 FEBS Lett. 275:151-158(1990).  
 [2]  
 SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 Champagne C.M.E., Maeda H., Takashiba S., van Dyke T.E.;  
 "Alternative splicing of diacylglycerol kinase alpha expressed in  
 human neutrophils.";  
 Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 [3]  
 MAPPING.  
 MEDLINE=94235961; PubMed=8180475;  
 Hart T.C., Champagne C., Zhou J., van Dyke T.E.;  
 "Assignment of the gene for diacylglycerol kinase (DAGK) to human  
 chromosome 12.";  
 Mamm. Genome 5:123-124(1994).  
 [4]  
 MAPPING.  
 MEDLINE=95048385; PubMed=7959783;  
 Hart T.C., Zhou J., Champagne C., van Dyke T.E., Rao P.N.,  
 Pettenati M.J.;  
 "Assignment of the human diacylglycerol kinase gene (DAGK) to 12q13.3  
 using fluorescence in situ hybridization analysis.";  
 Genomics 22:246-247(1994).  
 -1- FUNCTION: UPON CELL STIMULATION CONVERTS THE SECOND MESSENGER  
 OF PHOSPHATIDYLINOSITOLS AND ATTENUATING PROTEIN KINASE C  
 ACTIVITY.  
 -1- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL - ADP +  
 1,2-DIACYLGLYCEROL 3-PHOSPHATE.  
 -1- ENZYME REGULATED BY: STIMULATED BY CALCIUM AND PHOSPHATIDYLSELINE.  
 -1- SUBUNIT: MONOMER.  
 -1- TISSUE SPECIFICITY: LYMPHOCYTES AND OLIGODENDROGLIAL CELLS.  
 -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE  
 FAMILY.  
 -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 BINDING DOMAINS.  
 -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X62535; CAA44396.1; -  
 EMBL; AF064769; AAC34804.1; -  
 EMBL; AF064767; AAC34802.1; -  
 EMBL; AF064768; AAC34803.1; -  
 PIR; S12969; S12969.  
 MIM; 125855; -  
 InterPro; IPR000756; -

DR InterPro; IPR001206; -  
 DR InterPro; IPR002048; -  
 DR InterPro; IPR002219; -  
 DR Pfam; PF00609; DAGKa; 1.  
 DR Pfam; PF00781; DAGKc; 1.  
 DR Pfam; PF0130; DAG-PE-bind; 2.  
 DR Pfam; PF0036; eifhand; 2.  
 DR PRINTS; PR00008; DAGPEDOMAIN.  
 DR PROSITE; PS00479; DAG-PE\_BIND\_DOM\_1; 2.  
 DR PROSITE; PS00081; DAG-PE\_BIND\_DOM\_2; 2.  
 DR PROSITE; PS00018; EF\_HAND; 2.  
 DR Transferrase; Kinase; Calcium-binding; Phorbol-ester binding;  
 Multigene family.  
 FT CA\_BIND 123 134 SITE 1 (PROBABLE).  
 FT CA\_BIND 168 179 SITE 2 (PROBABLE).  
 FT DOMAIN 206 253 PHORBOL-ESTER AND DAG BINDING (BY  
 FT DOMAIN 270 319 SIMILARITY).  
 FT DOMAIN 374 500 CATALYTIC-A (POTENTIAL).  
 FT DOMAIN 520 701 CATALYTIC-B (POTENTIAL).  
 FT CONFLICT 339 339 L -> P (IN REF. 2).  
 FT CONFLICT 379 379 V -> L (IN AAC34802).  
 FT CONFLICT 385 385 S -> W (IN AAC34802).  
 FT CONFLICT 684 684 E -> G (IN AAC34803).  
 FT CONFLICT 699 699 V -> G (IN REF. 2).  
 FT CONFLICT 715 715 N -> K (IN REF. 2).  
 SQ SEQUENCE 735 AA; 82672 MW; ACAAD19DF4D510 CRC64;  
 Query Match 44.3%; Score 39; DB 1; Length 735;  
 Best Local Similarity 63.6%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 DWDTSALAPYL 11  
 DB 150 DWDVSELRPIL 160  
 RESULT 25  
 TPA\_HUMAN STANDARD; PRT; 562 AA.  
 AC P00750; Q15103; Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 21-JUL-1986 (Rel. 01, Last annotation update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TISSUE-TYPE PLASMINOGEN ACTIVATOR PRECURSOR (EC 3.4.21.68) (TPA)  
 DE (T-PA) (T-PLASMINOGEN ACTIVATOR) (ALTEPLASE) (RETEPLASE).  
 GN PLAT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=83115262; PubMed=6337343;  
 RA Pennica D., Holmes W.E., Kohr W.J., Harkins R.N., Vehar G.A.,  
 RA Ward C.A., Bennett W.F., Yelverton E., Seeburg P.H., Heyneker H.L.,  
 RA Goeddel D.V., Collen D.;  
 RT "Cloning and expression of human tissue-type plasminogen activator  
 cDNA in E. coli.";  
 RL Nature 301:214-221(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal lung;  
 RX MEDLINE=88262579; PubMed=3133640;  
 RA Sasaki H., Saito Y., Hayashi M., Otsuka K., Niwa M.;  
 RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA  
 from human fetal lung cells.";  
 RL Nucleic Acids Res. 16:5695-5695(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=88054470; PubMed=2824147;  
RA Reddy V.B., Garramone A.J., Sasak H., Wei C.-M., Watkins P., Galli J.,  
RA Hsiung N.;  
RT "Expression of human uterine tissue-type plasminogen activator in  
RT mouse cells using BPV vectors.";  
RL DNA 6:461-472(1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86196143; PubMed=3009482;  
RA Friezner Degen S.J., Rajput B., Reich E.;  
RT "The human tissue plasminogen activator gene.";  
RL J. Biol. Chem. 261:6972-6985(1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84298137; PubMed=6089198;  
RA Ny T., Elgh F., Lund B.;  
RT "The structure of the human tissue-type plasminogen activator gene:  
RT correlation of intron and exon structures to functional and  
RT structural domains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).  
RN [6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86284200; PubMed=3090401;  
RA Harris T.J., Patel T., Marston F.A., Little S., Emtage J.S.,  
RA Odenakker G., Volckaert G., Rombauts W., Billiau A., Somer P.;  
RT "Cloning of cDNA coding for human tissue-type plasminogen activator  
RT and its expression in Escherichia coli.";  
RL Mol. Biol. Med. 3:279-292(1986).  
RN [7]  
RP SEQUENCE OF 212-361 FROM N.A.  
RX MEDLINE=83169656; PubMed=6572897;  
RA Edlund T., Ny T., Raanby M., Heden L.-O., Palm G., Holmgren E.,  
RA Josephson S.;  
RT "Isolation of cDNA sequences coding for a part of human tissue  
RT plasminogen activator.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352(1983).  
RN [8]  
RP SEQUENCE OF 1-36 FROM N.A.  
RX MEDLINE=85289338; PubMed=3161893;  
RA Fisher R., Waller E.K., Grossi G., Thompson D., Tizard R.,  
RA Schleuning W.-D.;  
RT "Isolation and characterization of the human tissue-type plasminogen  
RT activator structural gene including its 5' flanking region.";  
RL J. Biol. Chem. 260:11223-11230(1985).  
RN [9]  
RP SEQUENCE OF 31-562 FROM N.A.  
RX MEDLINE=91291340; PubMed=1368681;  
RA Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.;  
RT "Purification and characterization of tissue plasminogen activator  
RT secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";  
RL Agric. Biol. Chem. 55:1225-1232(1991).  
RN [10]  
RP SEQUENCE OF 36-562.  
RC TISSUE-Melanoma;  
RX MEDLINE=85000468; PubMed=6433976;  
RA Pohl G., Kaellstroem M., Bergsdorf N., Wallen P., Joernvall H.;  
RT "Tissue plasminogen activator: peptide analyses confirm an indirectly  
RT derived amino acid sequence, identify the active site serine residue,  
RT establish glycosylation sites, and localize variant differences.";  
RL Biochemistry 23:3701-3707(1984).  
RN [11]  
RP SEQUENCE OF 33-52 AND 311-330.  
RC TISSUE-Melanoma;  
RX MEDLINE=83209620; PubMed=6682760;  
RA Wallen P., Pohl G., Bergsdorf N., Raanby M., Ny T., Joernvall H.;  
RT "Purification and characterization of a melanoma cell plasminogen  
RT activator.";  
RL Eur. J. Biochem. 132:681-686(1983).  
RN [12]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RC TISSUE-Umbilical vein;  
RX MEDLINE=90192129; PubMed=2107528;  
RA Siebert P.D., Fong K.;

RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from  
RT human endothelial cells.";  
RL Nucleic Acids Res. 18:1086-1086(1990).  
RN [13]  
RP STRUCTURE OF CARBOHYDRATES.  
RX MEDLINE=90092112; PubMed=2513186;  
RA Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.;  
RT "Carbohydrate structure of recombinant human uterine tissue  
RT plasminogen activator expressed in mouse epithelial cells.";  
RL Eur. J. Biochem. 186:273-286(1989).  
RN [14]  
RP CARBOHYDRATE-LINKAGE SITE THR-96.  
RX MEDLINE=91159408; PubMed=1900431;  
RA Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.;  
RT "Tissue plasminogen activator has an O-linked fucose attached to  
RT threonine-61 in the epidermal growth factor domain.";  
RL Biochemistry 30:2311-2314(1991).  
RN [15]  
RP DISULFIDE BONDS IN KRINGLE 2.  
RX MEDLINE=91244765; PubMed=1645336;  
RA Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.;  
RT "Disulfide pairing of the recombinant kringle-2 domain of tissue  
RT plasminogen activator produced in Escherichia coli.";  
RL J. Biol. Chem. 266:10070-10072(1991).  
RN [16]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.  
RX MEDLINE=96200985; PubMed=8613982;  
RA Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,  
RA Bode W.;  
RT "The 2.3 A crystal structure of the catalytic domain of recombinant  
RT two-chain human tissue-type plasminogen activator.";  
RL J. Mol. Biol. 258:117-135(1996).  
RN [17]  
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF CATALYTIC DOMAIN.  
RX MEDLINE=97449126; PubMed=9305622;  
RA Renatus M., Engh R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,  
RA Bode W.;  
RT "Lysine 156 promotes the anomalous proenzyme activity of tPA: X-ray  
RT crystal structure of single-chain human tPA.";  
RL EMBO J. 16:4797-4805(1997).  
RN [18]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.  
RX MEDLINE=92118803; PubMed=1310033;  
RA de Vos A., Ultsch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,  
RA Westbrook M.L., Kossiakof A.A.;  
RT "Crystal structure of the kringle 2 domain of tissue plasminogen  
RT activator at 2.4-A resolution.";  
RL Biochemistry 31:270-279(1992).  
RN [19]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=90122799; PubMed=2558718;  
RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
RT "1H NMR structural characterization of a recombinant kringle 2 domain  
RT from human tissue-type plasminogen activator.";  
RL Biochemistry 28:9350-9360(1989).  
RN [20]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=91200042; PubMed=1901789;  
RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR  
RT assignments and secondary structure.";  
RL Eur. J. Biochem. 197:155-165(1991).  
RN [21]  
RP STRUCTURE BY NMR OF KRINGLE 2.  
RX MEDLINE=92106329; PubMed=1762144;  
RA Byeon I.-J.L., Llinas M.;  
RT "Solution structure of the tissue-type plasminogen activator kringle  
RT 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic  
RT drug.";  
RL J. Mol. Biol. 222:1035-1051(1991).  
RN [22]  
RP STRUCTURE BY NMR OF 38-85.  
RX MEDLINE=92292163; PubMed=1602484;

RA Downing A.K., Driscoll P.C., Harvey T.S., Dudgeon T.J., Smith B.O.,  
RA Baron M., Campbell I.D.;  
RT "Solution structure of the fibrin binding finger domain of  
RT tissue-type plasminogen activator determined by 1H nuclear magnetic  
RT resonance.";  
RL J. Mol. Biol. 225:821-833(1992).  
RN [23]  
RP STRUCTURE BY NMR OF 36-126.  
RX MEDLINE=96027104; PubMed=7582899;  
RA Smith B.O., Downing A.K., Driscoll P.C., Dudgeon T.J., Campbell I.D.;  
RT "The solution structure and backbone dynamics of the fibronectin type  
RT I and epidermal growth factor-like pair of modules of tissue-type  
RT plasminogen activator.";  
RL Structure 3:823-833(1995).  
RN [23]  
CC -!- FUNCTION: CONVERTS THE ABUNDANT, BUT INACTIVE, ZMOGEN PLASMINOGEN  
CC TO PLASMIN BY HYDROLYZING A SINGLE ARG-VAL BOND IN PLASMINOGEN. BY  
CC CONTROLLING PLASMIN-MEDIATED PROTEOLYSIS, IT PLAYS AN IMPORTANT  
CC ROLE IN TISSUE REMODELING AND DEGRADATION, IN CELL MIGRATION AND  
CC MANY OTHER PHYSIOLOGICAL EVENTS.  
CC -!- CATALYTIC ACTIVITY: SPECIFIC CLEAVAGE OF ARG-|-VAL BOND IN  
CC PLASMINOGEN TO FORM PLASMIN.  
CC -!- SUBUNIT: BINDS TO FIBRIN WITH HIGH AFFINITY. THIS INTERACTION

Query Match 43.2%; Score 38; DB 1; Length 562;  
Best Local Similarity 44.4%; Pred. No. 74;  
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16  
DB 150 NWNSSALAKPKPYSGRRPD 167  
:::|||||:::

RESULT 26  
UGST\_ORYGL STANDARD; PRT; 609 AA.  
ID Q42968;  
AC 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GRANULE-BOUND GLYCOGEN [STARCH] SYNTHASE PRECURSOR (EC 2.4.1.11).  
GN WAXY.  
OS Oryza glaberrima (African rice).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Ehrhartoideae; Oryzeae;  
OC Oryza.  
OX NCBI\_TaxID=4538;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. GMS1;  
RX MEDLINE=92134825; PubMed=1685658;  
RA Umeda M., Ohtsubo H., Ohtsubo E.;  
RT "Diversification of the rice Waxy gene by insertion of mobile DNA  
RT elements into introns.";  
RL Jpn. J. Genet. 66:569-586(1991).  
CC -!- FUNCTION: REQUIRED FOR THE SYNTHESIS OF AMYLOSE IN ENDOSPERM. UDP +  
CC (1,4-BETA-D-GLUCOSYL)[N+1].  
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + (1,4-BETA-D-GLUCOSYL)[N] = UDP +  
CC (1,4-BETA-D-GLUCOSYL)[N+1].  
CC -!- PATHWAY: STARCH BIOSYNTHESIS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST, GRANULE-BOUND.  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; D10472; BAA01272.1;  
KW Glycogen biosynthesis; Transferase; Glycosyltransferase;  
KW Transit peptide; Chloroplast; Starch biosynthesis.

FT TRANSIT 1 77 CHLOROPLAST (BY SIMILARITY).  
FT CHAIN 78 609 GRANULE-BOUND GLYCOGEN [STARCH] SYNTHASE.  
FT BINDING 97 97 UDP-GLUCOSE (BY SIMILARITY).  
SQ SEQUENCE 609 AA; 56475 MW; C228BBFB9C407FA5 CRC64;  
  
Query Match 43.2%; Score 38; DB 1; Length 609;  
Best Local Similarity 63.6%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11  
DB 234 DWHGTPLASYL 244  
||| ||| ||| |||

RESULT 27  
UGST\_ORYSA STANDARD; PRT; 609 AA.  
ID P19395; Q43013;  
AC 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GRANULE-BOUND GLYCOGEN [STARCH] SYNTHASE PRECURSOR (EC 2.4.1.11).  
GN WAXY OR WX.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Ehrhartoideae; Oryzeae;  
OC Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Seed;  
RX MEDLINE=92322986; PubMed=13777969;  
RA Okagaki R.J.;  
RT "Nucleotide sequence of a long cDNA from the rice waxy gene.";  
RL Plant Mol. Biol. 19:513-516(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. JAPONICA TAICHUNG 65; TISSUE=Seedling;  
RA Hirano H.Y., Sano Y.;  
RT "Molecular characterization of the waxy locus of rice (Oryza  
RT sativa).";  
RL Plant Cell Physiol. 32:989-997(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. JAPONICA, AND CV. HANFENG;  
RX MEDLINE=91016948; PubMed=2216792;  
RA Wang Z.Y., Wu Z.L., Xing Y.Y., Zheng F.G., Guo X.L., Zhang W.G.,  
RA Hong M.M.;  
RT "Nucleotide sequence of rice waxy gene.";  
RL Nucleic Acids Res. 18:5898-5898(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. INDICA;  
RA Wang X.Q., Wang Z.Y., Hong M.M.;  
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 153-343 FROM N.A.  
RX MEDLINE=91200672; PubMed=2016064;  
RA Shimada H., Tada Y.;  
RT "Rapid isolation of a rice waxy sequence: a simple PCR method for the  
RT analysis of recombinant plasmids from intact Escherichia coli  
RT cells.";  
RL Gene 98:243-248(1991).  
CC -!- FUNCTION: REQUIRED FOR THE SYNTHESIS OF AMYLOSE IN ENDOSPERM.  
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + (1,4-ALPHA-D-GLUCOSYL)[N] = UDP  
CC + (1,4-ALPHA-D-GLUCOSYL)[N+1].  
CC -!- PATHWAY: STARCH BIOSYNTHESIS.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST, GRANULE-BOUND.  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X62134; CAA44065.1; -  
DR EMBL; X58228; CAA41186.1; -  
DR EMBL; X53694; CAA37732.1; -  
DR EMBL; X65183; CAA46294.1; -  
DR EMBL; M55039; AAA33918.1; -  
DR PIR; JQ0703; JQ0703.  
DR PIR; S22519; S22519.  
DR PIR; S30485; S30485.  
DR PIR; JQ2224; JQ2224.  
KW Glycogen biosynthesis; Transferase; Glycosyltransferase;  
KW Transist peptide; Chloroplast; Starch biosynthesis.  
FT TRANSIT 1 77 CHLOROPLAST.  
FT CHAIN 78 609 GRANULE-BOUND GLYCOGEN [STARCH] SYNTHASE.  
FT BINDING 97 97 UDP-GLUCOSE (BY SIMILARITY).  
FT CONFLICT 247 247 N -> T (IN REF. 5).  
FT CONFLICT 250 250 P -> T (IN REF. 5).  
FT CONFLICT 415 415 P -> S (IN REF. 4).  
SQ SEQUENCE 609 AA; 66476 MW; C225DBF6F12072C5 CRC64;

Query Match 43.2%; Score 38; DB 1; Length 609;  
Best Local Similarity 63.6%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11  
||| ||| |||  
DB 234 DWTGTPLASYL 244

RESULT 28  
AMVH\_SACDI  
ID AMVH\_SACDI STANDARD; PRT; 767 AA.  
AC P04065; Q92314;  
DT 01-NOV-1986 (Rel. 03, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLUCOAMYLASE S1 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-GLUCOSIDASE)  
DE (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE) (GAI).  
GN STAI1 OR DEX2 OR MAL5.  
OS Saccharomyces diastaticus (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=41870;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 60709;  
RX MEDLINE=85104778; PubMed=3918017;  
RA Yamashita I., Suzuki K., Fukui S.;  
RT "Nucleotide sequence of the extracellular glucoamylase gene STAI1 in  
RT the yeast Saccharomyces diastaticus.";  
RL J. Bacteriol. 161:567-573(1985).  
RN [2]  
RP SEQUENCE OF 1-54 FROM N.A.  
RA Yamashita I., Suzuki K., Sakuzo F.;  
RT "Proteolytic processing of glucoamylase in the yeast Saccharomyces  
RT cerevisiae.";  
RL Agric. Biol. Chem. 50:475-482(1986).  
RN [3]  
RP SEQUENCE OF 1-64 FROM N.A.  
RA Shima H., Inui M., Akada R., Yamashita I.;  
RT "Upstream regions of the yeast glucoamylase gene which are required  
RT for efficient transcription.";  
RL Agric. Biol. Chem. 53:749-755(1989).  
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-  
CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS  
CC WITH RELEASE OF BETA-D-GLUCOSE.

CC -!- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X02649; CAA26487.1; ALT\_INIT.

DR EMBL; D00428; BAA00332.1; -

DR PIR; A21896; ALBYG.

DR HSSP; P08017; LAYX.

DR InterPro; IPR000165; -

DR Pfam; PF00723; Glyco\_hydro\_15; 1.

DR PROSITE; PS00820; GLUCOAMYLASE; 1.

KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;

KW Signal; Multigene family.

FT SIGNAL 1 21

FT CHAIN 22 767 GLUCOAMYLASE S1.

FT DOMAIN 22 347 SER/THR-RICH.

FT DOMAIN 348 691 H SUBUNIT.

FT DOMAIN 692 767 Y SUBUNIT.

FT BINDING 455 455 SUBSTRATE (BY SIMILARITY).

FT ACT\_SITE 518 518 CATALYTIC BASE (BY SIMILARITY).

FT ACT\_SITE 521 521 GENERAL ACID CATALYST (BY SIMILARITY).

FT CARBOHYD 522 522 INTERACT WITH SUBSTRATES (BY SIMILARITY).

FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 767 AA; 82488 MW; A5F29E2427EDB593 CRC64;

Query Match 43.2%; Score 38; DB 1; Length 767;  
Best Local Similarity 46.7%; Pred. No. 1e+02;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQEE 16

||:||||:| |:

DB 333 WDSSALEEWLQROKK 347

RESULT 29

PGCB\_BOVIN

ID PGCB\_BOVIN STANDARD; PRT; 912 AA.

AC Q28062;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE BREVICAN CORE PROTEIN PRECURSOR.

GN BCAN.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Brain;

RX MEDLINE=94193597; PubMed=8144512;

RA Yamada H., Watanabe K., Shimonaka M., Yamaguchi Y.;

RT "Molecular cloning of brevicin, a novel brain proteoglycan of the

aggrecan/versican family.";

J. Biol. Chem. 269:10119-10126(1994).  
-!- FUNCTION: MAY PLAY A ROLE IN THE TERMINALLY DIFFERENTIATING AND THE ADULT NERVOUS SYSTEM DURING POSTNATAL DEVELOPMENT. COULD STABILIZE INTERACTIONS BETWEEN HA AND BRAIN PROTEOGLYCAN. SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY SIMILARITY).  
-!- TISSUE SPECIFICITY: BRAIN; EXPRESSED IN CEREBELLAR ASTROCYTES BUT NOT IN NEURONS.  
-!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE.  
-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
-!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
-!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
-!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
-!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
-!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: X75887; CAA53481.1; -.  
HSSP: P20693; 1HLJ.  
InterPro: IPR000436; -.  
InterPro: IPR000495; -.  
InterPro: IPR000538; -.  
InterPro: IPR000561; -.  
InterPro: IPR001304; -.  
InterPro: IPR003006; -.  
Pfam: PF00008; EGF\_1; 1.  
Pfam: PF00193; xlink; 1.  
Pfam: PF00047; ig; 1.  
Pfam: PF00059; lectin\_c; 1.  
Pfam: PF00084; sush1; 1.  
PROSITE: PS00022; EGF\_1; 1.  
PROSITE: PS01186; EGF\_2; 1.  
PROSITE: PS00290; IG\_MHC; 1.  
PROSITE: PS01241; LINK; 2.  
PROSITE: PS00615; C\_TYPE\_LLECTIN\_1; 1.  
PROSITE: PS00041; C\_TYPE\_LLECTIN\_2; 1.  
Glycoprotein: Hyaluronic acid; Proteoglycan; Lectin; Signal; Sush1; EGF-like domain; Repeat; Immunoglobulin domain.  
SIGNAL: 1 22  
CHAIN: 23 912  
DOMAIN: 32 158  
DOMAIN: 174 251  
DOMAIN: 272 353  
DOMAIN: 647 683  
DOMAIN: 683 811  
DOMAIN: 812 876  
DISULFID: 57 137  
DISULFID: 179 250  
DISULFID: 203 224  
DISULFID: 277 352  
DISULFID: 301 322  
DISULFID: 651 662  
DISULFID: 656 671  
DISULFID: 673 682  
DISULFID: 689 700  
DISULFID: 717 809  
DISULFID: 785 801  
DISULFID: 816 859  
DISULFID: 845 872  
CARBOHYD: 130 130  
CARBOHYD: 337 337  
SEQUENCE 912 AA: 99554 MW: 677B3EBIC688C4D7 CRC64;  
Query Match: 43.2%; Score 38; DB 1; Length 912;  
Best Local Similarity 37.5%; Pred. No. 1.2e+02;

Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
QY 1 DWDTSALAPYLGTQEE 16  
Db 433 EFEIQSIVPPLGSSE 448  
RESULT 30  
POLG\_EMCVB  
ID POLG\_EMCVB STANDARD; PRT; 2292 AA.  
AC P17593;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D (EC 2.7.7.48)]  
OS Encephalomyocarditis virus (strain emc-b nondiabetogenic).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Cardiovirus.  
OX NCBI\_TaxID=12105;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89243189; PubMed=2541543;  
RA Bae Y.S., Eun H.M., Yoon J.W.;  
RT "Genomic differences between the diabetogenic and nondiabetogenic variants of encephalomyocarditis virus."  
RL Virology 170:282-287(1989).  
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.  
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.  
CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: M22457; AAA43033.1; ALT\_SEQ.  
PIR: B31473; GNYEE.  
HSSP: P12296; LMEC.  
MEROPS: C03.009; -.  
DR MEROPS; U29.001; -.  
DR InterPro: IPR000605; -.  
DR InterPro: IPR001205; -.  
DR InterPro: IPR001676; -.  
DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam: PF00910; RNA\_helicase; 1.  
DR Pfam: PF00073; rhv; 3.  
KW Polyprotein; Coat protein; Core protein; Core protein; Transferase; RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate. PROPEP 1 67  
FT CHAIN 68 137  
FT CHAIN 138 393  
FT CHAIN 394 624  
FT CHAIN 625 901  
FT CHAIN 902 1058  
FT CHAIN 1059 1194  
FT CHAIN 1195 1519  
FT CHAIN 1520 1607  
FT CHAIN 1608 1627  
FT CHAIN 1628 1832  
FT CHAIN 1833 2292  
FT LIPID 68  
FT ACT\_SITE 1786 1786  
FT ACT\_SITE 1804 1804

```

SQ SEQUENCE 2292 AA; 255495 MW; 8540D0EB1437E8D4 CRC64;
Query Match 43.2%; Score 38; DB 1; Length 2292;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAPY 10
||:::| ||
Db 1959 DWESATLIPY 1968

RESULT 31
POLG EMCVD
ID POLG EMCVD STANDARD; PRT; 2292 AA.
AC P17594;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)].
OS Encephalomyocarditis virus (strain emc-d diabetogenic).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Cardiovirus.
OX NCBI_TaxID=12106;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89243189; PubMed=2541543;
RA Bae Y.S., Eun H.M., Yoon J.W.;
RT "Genomic differences between the diabetogenic and nondiabetogenic
RT variants of encephalomyocarditis virus.";
RL Virology 170:282-287(1989).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M22458; AAA43034.1; -.
CC PIR; A31473; GNNYED.
CC HSSP; P12296; IMEC.
CC MEROPS; C03.009; -.
CC MEROPS; U29.001; -.
CC InterPro; IPR000605; -.
CC InterPro; IPR001205; -.
CC InterPro; IPR001676; -.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC Pfam; PF00073; rhv; 3.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
FT PROPEP 1 67 LEADER PEPTIDE.
FT CHAIN 68 137 COAT PROTEIN VP4 (RHO).
FT CHAIN 138 393 COAT PROTEIN VP2 (BETA).
FT CHAIN 394 624 COAT PROTEIN VP3 (GAMMA).
FT CHAIN 625 901 COAT PROTEIN VP1 (ALPHA).
FT CHAIN 902 1058 CORE PROTEIN P2A (G).
FT CHAIN 1059 1194 CORE PROTEIN P2B (I).
FT CHAIN 1195 1519 CORE PROTEIN P2C (F).
FT CHAIN 1520 1607 CORE PROTEIN P3A.
FT CHAIN 1608 1627 GENOME-LINKED PROTEIN VPG (H).

SQ SEQUENCE 2292 AA; 255426 MW; F2B0627B0F444107 CRC64;
Query Match 43.2%; Score 38; DB 1; Length 2292;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAPY 10
||:::| ||
Db 1959 DWESATLIPY 1968

RESULT 32
MTX2 CAEEL
ID MTX2 CAEEL STANDARD; PRT; 260 AA.
AC P34599;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METAXIN 2 HOMOLOG.
GN ZC97.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaledon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL OUTER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE METAXIN FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L14714; AAA28180.1; -.
CC PIR; S44885; S44885.
CC WormPep; ZC97.1; CE00360.
KW Hypothetical protein; Mitochondrion; Outer membrane; Transport;
KW Protein transport.
SQ SEQUENCE 260 AA; 29752 MW; 4A9E46E937C3CFEE CRC64;
Query Match 42.6%; Score 37.5; DB 1; Length 260;
Best Local Similarity 53.3%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1 DW-DTSALAPYLGTO 14
|| | | | | | |
```



Db 22 DWEDVSLFTPYLNDQ 36

RESULT 33  
YNIF\_AZOBR STANDARD: PRT: 118 AA.

AC P25316;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DE 01-FEB-1996 (Rel. 33, Last annotation update)  
DE HYPOTHETICAL 12.3 KDA PROTEIN IN NIFK-NIFY INTERGENIC REGION.  
OS Azospirillum brasilense.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;  
OC Azospirillum.  
OX NCBI\_TaxID=192;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92370074; PubMed=1823284;  
RA Passaglia L.M.P.; Nunes C.P.; Zaha A.; Schrank I.S.;  
RT \*The nifHDK operon in the free-living nitrogen-fixing bacteria  
RT Azospirillum brasilense sequentially comprises genes H, D, K, an 353  
RT bp orf and gene Y.";  
RL Braz. J. Med. Biol. Res. 24:649-675(1991).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M64344; AAB02345.1; -  
CC PIR; S27476; S27476.  
CC Hypothetical protein; Nitrogen fixation.  
CC SEQUENCE 118 AA; 12323 MW; 54783E0F3F4D8D53 CRC64;  
CC -----  
CC Query Match 42.0%; Score 37; DB 1; Length 118;  
CC Best Local Similarity 77.8%; Pred. No. 24;  
CC Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
CC -----  
CC QY 1 DWDTSAALAP 9  
CC II III II  
CC 107 DWATSASAP 115  
CC -----  
CC Db 107 DWATSASAP 115

RESULT 34  
NHAA\_PSECL STANDARD: PRT: 199 AA.

AC P27764;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1993 (Rel. 26, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE NITRILE HYDRATASE SUBUNIT ALPHA (EC 4.2.1.84) (NITRILASE) (NHASE).  
DE NTHA.  
GN Pseudomonas chlororaphis (Pseudomonas fluorescens biotype D).  
OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas  
OX NCBI\_TaxID=333;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25 AND 163-177.  
RX STRAIN=B23;  
RX MEDLINE=91193202; PubMed=2013568;  
RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,  
RA Beppu T.;  
RT Cloning and characterization of genes responsible for metabolism of  
RT nitrile compounds from Pseudomonas chlororaphis B23.";  
RL J. Bacteriol. 173:2465-2472(1991).  
CC -!- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE  
CC COMPOUNDS TO THE CORRESPONDING AMIDES. INDUSTRIAL PRODUCTION OF  
CC ACRYLAMIDE IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.  
CC -!- CATALYTIC ACTIVITY: AN ALIPHATIC AMIDE - A NITRILE + H(2)O.

CC -!- COFACTOR: BINDS A NON-HEME IRON.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
CC -!- SIMILARITY: TO OTHER NITRILE HYDRATASES SUBUNIT ALPHA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D90216; BAA14245.1; -  
CC PIR; A42725; A42725.  
CC HSSP; P13448; LAHJ.  
KW Lyase; Iron.  
FT INIT\_MET 0  
FT METAL 104 104 IRON (BY SIMILARITY).  
FT METAL 107 107 IRON (BY SIMILARITY).  
FT METAL 108 108 IRON (BY SIMILARITY).  
FT METAL 109 109 IRON (BY SIMILARITY).  
SQ SEQUENCE 199 AA; 21986 MW; F580F6C428C251FE CRC64;  
CC -----  
CC Query Match 42.0%; Score 37; DB 1; Length 199;  
CC Best Local Similarity 70.0%; Pred. No. 39;  
CC Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
CC -----  
CC QY 2 WDTSAALAPYL 11  
CC IIII II  
CC 155 WDTSAESRYL 164  
CC -----  
CC RESULT 35  
SODM\_NOCAS STANDARD: PRT: 206 AA.

AC P53651;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).  
GN SODA OR SOD.  
OS Nocardia asteroides.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Nocardia.  
OX NCBI\_TaxID=1824;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GUH2;  
RX MEDLINE=96060854; PubMed=7590304;  
RA Alcendor D.J.; Chapman G.D.; Beaman B.L.;  
RT "Isolation, sequencing and expression of the superoxide dismutase-  
RT encoding gene (sod) of Nocardia asteroides strain GUH-2.";  
RL Gene 164:143-147(1995).  
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) - O(2) + H(2)O(2).  
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U02341; AAA91964.1; -  
CC HSSP; P17670; IIDS.  
DR InterPro; IPR001189; -  
DR Pfam; PF00081; scdfe; 1.  
DR PROSITE; PS00088; SOD\_MN; 1.



```
KW Oxidoreductase; Manganese.
FT INIT_MET 0 0
FT METAL 27 27
FT METAL 75 75
FT METAL 159 159
FT METAL 163 163
SQ SEQUENCE 206 AA; 22823 MW; FA2BCF27EB22CBCD CRC64;

Query Match 42.0%; Score 37; DB 1; Length 206;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWTSALAPYLGTQ 14
|:| ||| |:|
Db 9 DYDYSALEPHISGQ 22

RESULT 36
DGK1_LACAC STANDARD; PRT; 214 AA.
AC Q59483;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DEOXYADENOSINE KINASE (EC 2.7.1.76) (DADO KINASE) (DAK)
DE (DEOXYNUCLEOSIDE KINASE COMPLEX I S-COMPONENT).
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11506 / R-26;
RX MEDLINE=95204449; PubMed=7896798;
RA Ma G.T., Hong Y.S., Ives D.H.;
RT "Cloning and expression of the heterodimeric deoxyguanosine
RT kinase/deoxyadenosine kinase of Lactobacillus acidophilus R-26.";
RL J. Biol. Chem. 270:6595-6601(1995).
RN [2]
RP SEQUENCE OF 1-36.
RC STRAIN=ATCC 11506 / R-26;
RX MEDLINE=94227067; PubMed=8172906;
RA Ikeda S., Ma G.T., Ives D.H.;
RT "Heterodimeric deoxynucleoside kinases of Lactobacillus acidophilus
RT controlled by end-product inhibitors.";
RL Biochemistry 33:5328-5334(1994).
CC -!- FUNCTION: DGK/DAG PLAYS AN ESSENTIAL ROLE IN GENERATING THE
CC -!- DEOXYRIBONUCLEOTIDE PRECURSORS, DGTP AND DATP, FOR DNA METABOLISM.
CC -!- CATALYTIC ACTIVITY: ATP + DEOXYADENOSINE = ADP + DAMP.
CC -!- SUBUNIT: HETERODIMER OF A DEOXYADENOSINE (DAK) AND A
CC DEOXYGUANOSINE KINASE (DGK).
CC -!- SIMILARITY: BELONGS TO THE DCK/DGK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U01881; AAB09750.1; -.
DR InterPro; IPR002624; -.
DR Pfam; PF01712; dNK; 1.
KW Transferase; Kinase; ATP-binding.
FT INIT_MET 0 0
FT NP_BIND 7 14 ATP (POTENTIAL).
SQ SEQUENCE 214 AA; 24565 MW; 64599B958728FC10 CRC64;

Query Match 42.0%; Score 37; DB 1; Length 214;
```

```
Best Local Similarity 72.7%; Pred. No. 42;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 TSALAPYLGTQ 14
|:| || |:|
Db 17 TSLLAEHLGTQ 27

RESULT 37
YP97_CAEEL STANDARD; PRT; 283 AA.
AC Q09245;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 33.3 KDA PROTEIN C28H8.7 IN CHROMOSOME III.
GN C28H8.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Miller N., Waterston R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U20861; AAA62295.1; -.
DR WormPep; C28H8.7; CE01829.
KW Hypothetical protein.
SQ SEQUENCE 283 AA; 33295 MW; CD8941ABE66CE9F8 CRC64;

Query Match 42.0%; Score 37; DB 1; Length 283;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQE 15
|:| ||| |:|
Db 159 WDYGSLANYLNQSE 172

RESULT 38
THTR_RAT STANDARD; PRT; 295 AA.
AC P24329;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE THIOSULFATE SULFURTRANSFERASE (EC 2.8.1.1) (RHODANESE) (FRAGMENT).
GN TST.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=91207296; PubMed=2018478;
RA Weiland K.L., Dooley T.P.;
RT "Molecular cloning, sequencing and characterization of cdna to rat
RT liver rhodanese, a thiosulphate sulphurtransferase.";
RL Biochem. J. 275:227-231(1991).
RN [2]
RP MUTAGENESIS.
```

RC TISSUE=Liver; MEDLINE=95332330; PubMed=7608189; GN  
RX Nagahara N., Okazaki T., Nishino T.; OS  
RA "Cytosolic mercaptopyruvate sulfoxidation is evolutionarily OC  
RT related to mitochondrial rhodanese. Striking similarity in active site OC  
RT amino acid sequence and the increase in the mercaptopyruvate OX  
RT sulfoxidation activity of rhodanese by site-directed RN  
RT mutagenesis." J. Biol. Chem. 270:16230-16235(1995). RP  
RL J. Biol. Chem. 270:16230-16235(1995). RX  
CC J. Biol. Chem. 270:16230-16235(1995). RA  
CC FUNCTION: INVOLVED IN THE FORMATION OF IRON-SULFUR COMPLEXES, RT  
CC CYANIDE DETOXIFICATION OR MODIFICATION OF SULFUR-CONTAINING RT  
CC ENZYMES. OTHER THIOL COMPOUNDS, BESIDES CYANIDE, CAN ACT AS SULFUR RL  
CC ION ACCEPTORS. ALSO HAS WEAK MERCAPTOPYRUVATE SULFOTRANSFERASE RN  
CC (MST) ACTIVITY. RP  
CC CATALYTIC ACTIVITY: THIOSULFATE + CYANIDE = SULFITE + THIOCYANATE. RC  
CC SUBUNIT: MONOMER. RX  
CC SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX. RA  
CC TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES. RT  
CC DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR RT  
CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER, RL  
CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT. J. Biol. Chem. 253:8102-8108(1978).  
CC SIMILARITY: BELONGS TO THE RHODANESE FAMILY. [3]  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)). [4]  
CC EMBL; X56228; CAA39677.1; --  
CC PIR; S15081; S15081.  
CC HSSP; P00586; IORB.  
CC InterPro; IPR001307; --  
CC InterPro; IPR001763; --  
CC Pfam; PF00581; Rhodanese; 2.  
CC PROSITE; PS00380; RHODANESE\_1; 1.  
CC PROSITE; PS00683; RHODANESE\_2; 1.  
CC TRANSFERASE; Mitochondrion.  
CC NON\_TER 1 141 A DOMAIN.  
CC DOMAIN 1 141 HINGE.  
CC DOMAIN 142 157 B DOMAIN.  
CC DOMAIN 158 295 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY  
CC ACT\_SITE 185 SIMILARITY).  
CC ACT\_SITE 246  
CC ACT\_SITE 247  
CC ACT\_SITE 248  
CC MUTAGEN 247  
CC MUTAGEN 248  
CC SEQUENCE 295 AA; 33176 MW; 24C55B35690934E1 CRC64;  
Query Match 42.0%; Score 37; DB 1; Length 295;  
Best Local Similarity 57.1%; Pred No. 58;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 3 DTSALAPYLGTQEE 16  
Db 179 DSRAQGRYLGTPQE 192  
RESULT 39  
ID THTR\_BOVIN STANDARD; PRT; 296 AA.  
AC P00586;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE THIOSULFATE SULFOTRANSFERASE (EC 2.8.1.1) (RHODANESE).

GN TST. OS  
OS Bos taurus (Bovine). OC  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; OC  
OC Bovidae; Bovinae; Bos. OX  
OX NCBI\_TaxID=9913; RN  
RN SEQUENCE FROM N.A. RP  
RX MEDLINE=91161544; PubMed=2002017; RX  
RA Miller D.M., Delgado R., Chirgwin J.M., Hardies S.C., Horowitz P.M.; RA  
RT "Expression of cloned bovine adrenal rhodanese." J. Biol. Chem. 266:4686-4691(1991). RT  
RL J. Biol. Chem. 266:4686-4691(1991). RL  
RN SEQUENCE OF 1-294. RP  
RP TISSUE=Liver; RC  
RX MEDLINE=79048424; PubMed=711737; RX  
RA Russell J., Weng L., Keim P.S., Heinrichson R.L.; RA  
RT "The covalent structure of bovine liver rhodanese. Isolation and RT  
RT partial structural analysis of cyanogen bromide fragments and the RT  
RT complete sequence of the enzyme." J. Biol. Chem. 253:8102-8108(1978). RT  
RL J. Biol. Chem. 253:8102-8108(1978). RL  
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS). [3]  
RP MEDLINE=79007483; PubMed=691057; RP  
RX Ploegman J.H., Drent G., Kalk K.H., Hol W.G.J.; RX  
RA "Structure of bovine liver rhodanese. I. Structure determination at RT  
RT 2.5-A resolution and a comparison of the conformation and sequence of RT  
RT its two domains." J. Mol. Biol. 123:557-594(1978). RT  
RL J. Mol. Biol. 123:557-594(1978). RL  
RN X-RAY CRYSTALLOGRAPHY (1.36 ANGSTROMS). [4]  
RP MEDLINE=98437562; PubMed=9761843; RP  
RX Gliubich F., Berni R., Colapletto M., Barba L., Zanotti G.; RX  
RT "Structure of sulfur-substituted rhodanese at 1.36-A resolution." Acta Crystallogr. D 54:481-486(1998). RT  
RN ACTIVE SITE. [5]  
RP MEDLINE=79048425; PubMed=711738; RP  
RX Weng L., Heinrichson R.L., Westley J.; RX  
RA "Active site cysteinyl and arginyl residues of rhodanese. A novel RT  
RT formation of disulfide bonds in the active site promoted by RT  
RT phenylglyoxal." J. Biol. Chem. 253:8109-8119(1978). RT  
RL J. Biol. Chem. 253:8109-8119(1978). RL  
RN MUTAGENESIS OF ARG-186 AND LYS-249. [6]  
RP MEDLINE=94179198; PubMed=8132546; RP  
RX Luo G.-X., Horowitz P.M.; RX  
RT "The sulfoxidation activity and structure of rhodanese are RT  
RT affected by site-directed replacement of Arg-186 or Lys-249." J. Biol. Chem. 269:8220-8225(1994). RT  
RL J. Biol. Chem. 269:8220-8225(1994). RL  
CC FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE  
CC DETOXIFICATION. BINDS MOLECULAR OXYGEN AND SULFUR.  
CC CATALYTIC ACTIVITY: THIOSULFATE + CYANIDE = SULFITE + THIOCYANATE.  
CC SUBUNIT: MONOMER.  
CC SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
CC TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES.  
CC DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
CC SIMILARITY: BELONGS TO THE RHODANESE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M58561; AAA30753.1; --  
CC PIR; A00727; ROBO.  
CC PIR; A23704; A23704.  
CC PDB; 1RHD; 27-JAN-84.

DR PDB; 1RHS; 21-JAN-98.  
DR PDB; 2ORA; 01-AUG-96.  
DR PDB; 1ORB; 15-OCT-95.  
DR PDB; 1BOH; 27-APR-99.  
DR PDB; 1BOI; 27-APR-99.  
DR InterPro; IPR001307; -.  
DR InterPro; IPR001763; -.  
DR Pfam; PF00581; Rhodanese; 2.  
DR PROSITE; PS00380; RHODANESE\_1; 1.  
DR PROSITE; PS00683; RHODANESE\_2; 1.  
KW Transferase; Mitochondrion; 3D-structure.  
FT INIT\_MET 0 0  
FT DOMAIN 1 142 A DOMAIN.  
FT DOMAIN 143 158 HINGE.  
FT DOMAIN 159 296 B DOMAIN.  
FT ACT\_SITE 186 186 MAY PLAY A ROLE IN SUBSTRATE BINDING.  
FT ACT\_SITE 247 247  
FT ACT\_SITE 248 248  
FT ACT\_SITE 249 249  
FT VARIANT 1 2  
FT MUTAGEN 186 186  
FT MUTAGEN 249 249  
FT CONFLICT 99 99  
FT CONFLICT 214 214  
FT CONFLICT 219 219  
FT STRAND 9 10  
FT HELIX 12 21  
FT TURN 22 22  
FT STRAND 24 24  
FT TURN 25 27  
FT STRAND 28 32  
FT TURN 38 40  
FT HELIX 43 47  
FT TURN 48 49  
FT STRAND 51 51  
FT TURN 53 54  
FT STRAND 56 57  
FT TURN 60 61  
FT TURN 66 67  
FT HELIX 77 87  
FT TURN 91 92  
FT STRAND 94 97  
FT TURN 102 103  
FT HELIX 108 117  
FT TURN 118 118  
FT STRAND 123 126  
FT TURN 127 128  
FT HELIX 129 136  
FT TURN 137 137  
FT STRAND 141 141  
FT TURN 158 159  
FT STRAND 161 162  
FT TURN 164 165  
FT HELIX 166 173  
FT STRAND 177 180  
FT HELIX 184 188  
FT TURN 189 189  
FT STRAND 203 203  
FT TURN 205 206  
FT STRAND 208 209  
FT TURN 212 215  
FT STRAND 216 216  
FT TURN 218 219  
FT STRAND 222 222  
FT HELIX 225 234  
FT TURN 235 236  
FT TURN 239 240  
FT STRAND 243 246  
FT TURN 252 253  
FT HELIX 254 263

FT TURN 264 264  
FT TURN 266 267  
FT STRAND 269 271  
FT TURN 272 274  
FT HELIX 275 281  
FT HELIX 284 286  
FT STRAND 288 288  
SQ SEQUENCE 296 AA; 33164 MW; C8769696FA6AC111 CRC64;  
  
Query Match 42.0%; Score 37; DB 1; Length 296;  
Best Local Similarity 57.1%; Pred. No. 58;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 3 DTSALAPYLGTQEE 16  
I: I I I I I I  
Db 180 DSRAGRYLGTQPE 193  
  
RESULT 40  
THTR\_MOUSE  
ID THTR\_MOUSE STANDARD; PRT; 296 AA.  
AC P52196;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE THIOSULFATE SULFURTRANSFERASE (EC 2.8.1.1) (RHODANESE).  
GN TST.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=Liver;  
RX MEDLINE=96074596; PubMed=7488186;  
RA Dooley T.P., Nair S.K., Garcia R.E., Courtney B.C.;  
RT "Mouse rhodanese gene (Tst): cDNA cloning, sequencing, and  
RT recombinant protein expression."  
RL Biochem. Biophys. Res. Commun. 216:1101-1109(1995).  
CC -!- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE  
CC DETOXIFICATION.  
CC -!- CATALYTIC ACTIVITY: THIOSULFATE + CYANIDE = SULFITE + THIOCYANATE.  
CC -!- SUBUNIT: MONOMER.  
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
CC -!- TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES.  
CC -!- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
CC -!- SIMILARITY: BELONGS TO THE RHODANESE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U35741; AAC52342.1; -.  
DR HSSP; P00586; 1RHD.  
DR SWISS-2DPAGE; P52196; MOUSE.  
DR MGD; MGI:98852; Tst.  
DR InterPro; IPR001307; -.  
DR InterPro; IPR001763; -.  
DR Pfam; PF00581; Rhodanese; 2.  
DR PROSITE; PS00380; RHODANESE\_1; 1.  
DR PROSITE; PS00683; RHODANESE\_2; 1.  
KW Transferase; Mitochondrion.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT DOMAIN 1 142 A DOMAIN.  
FT DOMAIN 143 158 HINGE.  
FT DOMAIN 159 296 B DOMAIN.

```
FT ACT_SITE 186 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY
FT ACT_SITE 247 SIMILARITY).
FT ACT_SITE 248 BY SIMILARITY.
FT ACT_SITE 249 SUBSTRATE (THIOSULFATE) BINDING
FT ACT_SITE 249 (BY SIMILARITY).
FT SEQUENCE 296 AA; 33334 MW; 82089D880F9AE55A CRC64;
SQ

Query Match 42.0%; Score 37; DB 1; Length 296;
Best Local Similarity 57.1%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 DTSALAPYLGTQEE 16
DB 180 DSRAQRYLGTQPE 193

RESULT 41
VG03_HSV11 STANDARD; PRT; 302 AA.
AC Q00115;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE HYPOTHETICAL GENE 3 PROTEIN.
GN 3.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC unclassified Herpesviridae.
OC NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUBURN 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RL "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M75136; AAA88184.1; -
CC EMBL; M75136; AAA88106.1; -
CC PIR; D36786; D36786.
CC Hypothetical protein.
SQ SEQUENCE 302 AA; 32282 MW; B03FE6B3180F19FD CRC64;

Query Match 42.0%; Score 37; DB 1; Length 302;
Best Local Similarity 42.9%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLGTQ 14
DB 84 DWDSPASPPVTGAE 97

RESULT 42
HPRK_TREPA STANDARD; PRT; 319 AA.
AC O83600;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HPR(SER) KINASE/PHOSPHATASE (EC 2.7.1.-) (EC 3.1.3.-).
GN HPRK OR PTSK OR TP0591.
```

```
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Peterson J.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Utterback T.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -----
CC FUNCTION: THIS KINASE/PHOSPHATASE REGULATES CARBOHYDRATE UPTAKE
CC AND METABOLISM BY PHOSPHORYLATING/DEPHOSPHORYLATING SER-46 OF THE
CC PHOSPHORYL CARRIER PROTEIN (HPR) OF THE PHOSPHOENOLPYRUVATE-
CC DEPENDENT SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS) IN RESPONSE TO
CC CYTOPLASMIC METABOLITE LEVELS AS A DEVICE FOR THE MODULATION OF
CC CARBON CATABOLITE REPRESSION (BY SIMILARITY).
CC -----
CC SIMILARITY: BELONGS TO THE PTSK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001234; AAC65566.1; -
CC TIGR; TP0591; -
CC Multifunctional enzyme; Transferase; Serine/threonine-protein kinase;
KW ATP-binding; Hydrolase.
FT NP_BIND 152 159
SQ SEQUENCE 319 AA; 35548 MW; 5ADF40485574B12D CRC64;

Query Match 42.0%; Score 37; DB 1; Length 319;
Best Local Similarity 46.7%; Pred. No. 62;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLGTQE 15
DB 234 EWNSSKAYDRLGTQE 248

RESULT 43
PKX1_HUMAN STANDARD; PRT; 358 AA.
AC P51817;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTEIN KINASE PKX1 (EC 2.7.1.-).
GN PRKX OR PKX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RC MEDLINE=95360006; PubMed=76333447;
RA Klink A., Schiebel K., Winkelmann M., Rao E., Horsthemke B.,
RA Lueddecke H.-J., Claussen U., Scherer G., Rappold G.;
RT "The human protein kinase gene PKX1 on Xp22.3 displays xp/yp homology
RT and is a site of chromosomal instability.";
RL Hum. Mol. Genet. 4:869-878(1995).
CC -----
CC TISSUE SPECIFICITY: HIGH LEVELS IN ADULT AND FETAL BRAIN, KIDNEY
```



CC AND LUNG; LOW LEVELS IN ADULT PLACENTA, HEART, LIVER, SKELETAL  
CC MUSCLE, PANCREAS, AND FETAL LIVER.  
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-  
CC PROTEIN KINASES. TO CAMP-DEPENDENT PROTEIN KINASES. MOST SIMILAR  
CC TO DROSOPHILA DC2.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X85545; CAA59733.1; -.  
DR HSP; P05132; 2CPK.  
DR MIM; 300083; -.  
DR InterPro; IPR000719; -.  
DR InterPro; IPR000961; -.  
DR InterPro; IPR002290; -.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00433; pkinase\_C; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
FT DOMAIN 49 303 PROTEIN KINASE.  
FT NP\_BIND 55 63 ATP (BY SIMILARITY).  
FT BINDING 78 78 ATP (BY SIMILARITY).  
FT ACT\_SITE 172 172 BY SIMILARITY.  
SQ SEQUENCE 358 AA; 40896 MW; B603834E3541CB56 CRC64;

Query Match 42.0%; Score 37; DB 1; Length 358;  
Best Local Similarity 66.7%; Pred. No. 70;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAP 9  
|||||I|I|  
Db 340 DWDTAAPVP 348

RESULT 44  
METX\_NEIMB STANDARD; PRT; 379 AA.  
AC Q9JZQ5;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.31) (HOMOSERINE O-TRANS-  
DE ACETYLASE) (HOMOSERINE TRANSACETYLASE) (HTA).  
GN METX OR NMB0940.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / SEROGROUP B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,  
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
RT MC58.";  
RL Science 287:1809-1815(2000).  
CC -!- CATALYTIC ACTIVITY: ACETYL-COA + L-HOMOSERINE = COA + O-ACETYL-L-  
CC HOMOSERINE.  
CC -!- PATHWAY: BIOSYNTHESIS OF METHIONINE; HTA VARIANT; FIRST STEP.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE AB HYDROLASE FAMILY; HTA SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE002445; AAF41346.1; -.  
DR TIGR; NMB0940; -.  
DR InterPro; IPR000073; -.  
DR Pfam; PF00561; abhydrolase; 1.  
KW Methionine biosynthesis; Transferase; Acyltransferase.  
FT ACT\_SITE 154 154 POTENTIAL.  
FT ACT\_SITE 352 352 POTENTIAL.  
SQ SEQUENCE 379 AA; 42173 MW; D56AC364101EE8EE CRC64;

Query Match 42.0%; Score 37; DB 1; Length 379;  
Best Local Similarity 47.1%; Pred. No. 74;  
Matches 8; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

QY 1 DW--DTSALAPYLGTOE 15  
|| :||| ||| :|  
Db 130 DWKSOAALADYLGIEQ 146

RESULT 45  
AMYG\_YEAST STANDARD; PRT; 549 AA.  
AC P08019;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE GLUCOAMYLASE, INTRACELLULAR SPORULATION-SPECIFIC (EC 3.2.1.3) (GLUCAN  
DE 1,4-ALPHA-GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).  
GN SGA1 OR SGA OR YIL099W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87194600; PubMed=3106330;  
RA Yamashita I., Nakamura M., Fukui S.;  
RT "Gene fusion is a possible mechanism underlying the evolution of  
RT STA1.";  
RL J. Bacteriol. 169:2142-2149(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Barrell B.G., Badcock K., Bankier A.T., Bowan S., Brown D.,  
RA Churcher C.M., Connor R., Copestake T., Dear S., Devlin K., Fraser A.,  
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
RA Walsh S.V., Whitehead S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-  
CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS  
CC WITH RELEASE OF BETA-D-GLUCOSE.  
CC -!- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC -----  
DR EMBL; Z38125; CAA86282.1; -;  
DR EMBL; M16166; AAA35042.1; -;  
DR PIR; C26877; C26877;  
DR PIR; S48474; S48474;  
DR HSSP; P08017; LAYX;  
DR SGD; S0001361; SGAL;  
DR InterPro; IPR000165;  
DR Pfam; PF00723; Glyco\_hydro\_15; 1;  
DR PRINTS; PR00736; GLHYDRLASE15;  
DR PROSITE; PS00820; GLUCOAMYLASE; 1;  
KW Hydrolase; Glycosidase; Polysaccharide degradation; Sporulation.  
FT BINDING 198 198 SUBSTRATE (BY SIMILARITY).  
FT ACT\_SITE 261 261 CATALYTIC BASE (BY SIMILARITY).  
FT ACT\_SITE 264 264 GENERAL ACID CATALYST (BY SIMILARITY).  
FT ACT\_SITE 265 265 INTERACT WITH SUBSTRATES (BY SIMILARITY).  
FT CONFLICT 504 549 HVGTDGELSEQFNKYTGFMQGAQHLTWSYTSFWDAYQIRQE  
FT VLSL -> TWEQTGN (IN REF. 1).  
SQ SEQUENCE 549 AA; 61463 MW; 6351E84F2CF4AB77 CRC64;

Query Match 42.0%; Score 37; DB 1; Length 549;  
Best Local Similarity 46.7%; Pred. NO. 1.1e+02;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 WDTSALAPYLGTQEE 16  
Db 76 WDSSALEEWLQGQKK 90

Search completed: June 28, 2001, 11:57:27  
Job time: 268 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:32 ; Search time 72.61 Seconds  
(without alignments)  
29.154 Million cell updates/sec

Title: US-09-439-313-562  
Perfect score: 88  
Sequence: 1 DWDTSALAPYLGTQEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SPTREMBL\_16:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB	ID	Description
1	46	52.3	491	2	P94289		P94289 bacillus ci
2	45	51.1	685	2	Q9K1Q0		Q9k1q0 neisseria m
3	45	51.1	685	2	Q9JWP0		Q9jwp0 neisseria m
4	45	51.1	831	2	O50076		O50076 clostridium
5	45	51.1	847	2	Q9I4U2		Q9i4u2 pseudomonas
6	43	48.9	207	2	Q9F9R1		Q9f9r1 mycobacteri
7	43	48.9	275	2	Q9I567		Q9i567 pseudomonas
8	43	48.9	402	2	Q55118		Q55118 synechocyst
9	43	48.9	459	2	Q9X7Q9		Q9x7q9 streptomyc
10	42	47.7	227	3	P79022		P79022 candida sp.
11	42	47.7	490	2	Q9PLR4		Q9plr4 chlamydia m
12	42	47.7	563	14	Q9N2J9		Q9n2j9 herv-h/envv6
13	41.5	47.2	1044	2	Q9F4D6		Q9f4d6 bifidobacte
14	41	46.6	217	1	O58748		O58748 pyrococcus
15	41	46.6	332	2	Q9S3X3		Q9s3x3 streptomyc
16	41	46.6	343	2	O85977		O85977 sphingomona
17	41	46.6	343	2	Q92306		Q923u6 sphingomona
18	41	46.6	359	6	Q28450		Q28450 loxodonta a
19	41	46.6	410	6	Q28556		Q28556 orycteropus

20	41	46.6	427	1	Q9Y9T8	Q9y9t8 aeropyrum p
21	41	46.6	2487	6	Q9N1T0	Q9n1t0 ornithorhyn
22	41	46.6	9477	2	Q9L4X3	Q9l4x3 streptomyc
23	40	45.5	258	5	O16266	O16266 caenorhabdi
24	40	45.5	294	2	Q9I1R6	Q9i1r6 pseudomonas
25	40	45.5	392	10	Q9M6J3	Q9m6j3 aquilegia s
26	40	45.5	440	10	Q9LHI4	Q9lhi4 arabisdopsis
27	40	45.5	482	2	Q59326	Q59326 clostridium
28	40	45.5	490	10	Q9FFL1	Q9ffl1 arabisdopsis
29	40	45.5	732	10	Q9XE45	Q9xe45 arabisdopsis
30	40	45.5	957	10	Q9SDA6	Q9sda6 arabisdopsis
31	40	45.5	1069	5	Q9VSC9	Q9vsc9 drosophila
32	40	45.5	1265	2	P72316	P72316 rhodospiril
33	40	45.5	2047	4	O15019	O15019 homo sapien
34	39.5	44.9	575	1	Q9HSG5	Q9hsg5 halobacteri
35	39	44.3	81	2	O53813	O53813 mycobacteri
36	39	44.3	84	10	Q9LGN9	Q9lgn9 oryza sativ
37	39	44.3	175	2	O53803	O53803 mycobacteri
38	39	44.3	179	2	P78203	P78203 escherichia
39	39	44.3	196	13	P79827	P79827 oncorhynchu
40	39	44.3	209	2	Q9L0K0	Q9l0k0 streptomyc
41	39	44.3	282	6	Q9GJT2	Q9gjt2 sus scrofa
42	39	44.3	326	2	O54397	O54397 rhodococcus
43	39	44.3	337	5	O16739	O16739 caenorhabdi
44	39	44.3	346	1	O58813	O58813 pyrococcus
45	39	44.3	347	1	Q9HP38	Q9hp38 halobacteri
46	39	44.3	387	2	O55481	O55481 synechocyst
47	39	44.3	464	6	O46520	O46520 sus scrofa
48	39	44.3	479	10	Q9LH39	Q9lh39 arabisdopsis
49	39	44.3	504	1	O29585	O29585 archaeoglob
50	39	44.3	580	2	Q9I134	Q9i134 pseudomonas
51	39	44.3	677	2	Q9HYC7	Q9hyc7 pseudomonas
52	39	44.3	702	2	Q9PFV8	Q9pfv8 xylella fas
53	39	44.3	731	2	Q9KT69	Q9kt69 vibrio chol
54	39	44.3	948	2	Q9HZE4	Q9hze4 pseudomonas
55	39	44.3	1020	4	O94928	O94928 homo sapien
56	39	44.3	2293	14	Q83422	Q83422 mengo encep
57	39	44.3	2316	2	Q9FDJ9	Q9fdj9 bacteroides
58	39	44.3	4124	5	O97218	O97218 leishmania
59	38	43.2	178	10	Q9SDN1	Q9sdn1 prunus dulc
60	38	43.2	198	14	Q83368	Q83368 murine leuk
61	38	43.2	205	5	Q9GPR2	Q9gpr2 dictyosteli
62	38	43.2	215	1	Q9HJ08	Q9hj08 thermoplasm
63	38	43.2	224	10	O65523	O65523 arabisdopsis
64	38	43.2	247	2	P73809	P73809 synechocyst
65	38	43.2	256	10	O22451	O22451 zea mays (m

ALIGNMENTS

RESULT	1
P94289	
ID	P94289
AC	P94289;
DT	01-MAY-1997 (TrEMBLrel. 03, Created)
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE	CHITINASE C PRECURSOR.
OS	Bacillus circulans.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1397;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=WL-12;
RA	Alam M.;
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=WL-12;
RA	Alam M., Nikaidou N., Tanaka H., Watanabe T.;
RT	*Cloning and sequencing of chiC gene of Bacillus circulans WL-12 and

RT relationship of its product to some other chitinases and chitinase-  
like proteins.";  
RL J. Ferment. Bioeng. 80:454-461(1995).  
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
HYDROLASES).  
DR EMBL; D89568; BAA13974.1; --  
DR HSSP; P07254; 1CTN.  
DR InterPro; IPR001223; --  
DR InterPro; IPR001579; --  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase; Signal.  
FT SIGNAL 1 41  
FT CHAIN 42 491  
FT POTENTIAL.  
FT POTENTIAL.  
SQ SEQUENCE 491 AA; 53447 MW; 31A98122AE9BCBEC CRC64;  
  
Query Match 52.3%; Score 46; DB 2; Length 491;  
Best Local Similarity 80.0%; Pred. No. 9.7;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 WDTSALAPYL 11  
Db 318 WDSSAQAPYL 327  
  
RESULT 2  
Q9K1Q0 PRELIMINARY; PRT; 685 AA.  
ID Q9K1Q0;  
AC Q9K1Q0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE METHIONYL-TRNA SYNTHETASE.  
GN NMB0030.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / SEROGROUP B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,  
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
MC58".  
RT Science 287:1809-1815(2000).  
RL EMBL; AE002361; AAF40501.1; --  
DR TIGR; NMB0030; --  
DR InterPro; IPR001412; --  
DR InterPro; IPR002304; --  
DR InterPro; IPR002547; --  
DR Pfam; PF01588; tRNA\_bind; 1.  
DR PROSITE; PR01041; TRNASYNTHMET.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
KW Aminoacyl-tRNA synthetase.  
SQ SEQUENCE 685 AA; 76852 MW; 9C0F39507B1FD50A CRC64;

Query Match 51.1%; Score 45; DB 2; Length 685;  
Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 4; Indels 4; Gaps 0;  
  
QY 1 DWDTLSALAPYL 12  
Db 237 DWDISRDAPYFG 248

RESULT 3  
Q9JWP0 PRELIMINARY; PRT; 685 AA.  
ID Q9JWP0;  
AC Q9JWP0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10).  
GN METG OR NMA0275.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;  
RX MEDLINE=20222556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrell B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
meningitidis Z2491".  
RT Nature 404:502-506(2000).  
RL EMBL; AL162752; CAB83583.1; --  
DR InterPro; IPR001412; --  
DR InterPro; IPR002304; --  
DR InterPro; IPR002547; --  
DR Pfam; PF01588; tRNA\_bind; 1.  
DR PRINTS; PR01041; TRNASYNTHMET.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
KW Aminoacyl-tRNA synthetase; Ligase.  
SQ SEQUENCE 685 AA; 76984 MW; 07FD45915ED3BEF3 CRC64;  
  
Query Match 51.1%; Score 45; DB 2; Length 685;  
Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 4; Indels 4; Gaps 0;  
  
QY 1 DWDTLSALAPYL 12  
Db 237 DWDISRDAPYFG 248  
  
RESULT 4  
O50076 PRELIMINARY; PRT; 831 AA.  
ID O50076;  
AC O50076; 1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-JUN-1998 (TREMBLrel. 16, Last annotation update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CHITINASE B (EC 3.2.1.14) (CHITODEXTRINASE) (1,4-BETA-POLY-N-  
DE ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).  
GN ChIB.  
OS Clostridium paraputrificum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=29363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M21;  
RX MEDLINE=98053843; PubMed=9393694;  
RA Ohmiya K., Sakka K., Kimura T., Karita S., Morimoto K.;  
RT "Cloning, sequencing, and expression of the gene encoding Clostridium  
paraputrificum chitinase ChIB and analysis of the functions of novel  
RT cadherin-like domains and a chitin-binding domain.";  
RL J. Bacteriol. 179:7306-7314(1997).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-  
CC ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
CC EMBL; AB001874; BAA23796.1; --  
DR HSSP; P07254; 1CTN.



```
DR InterPro; IPR001223; -.
DR InterPro; IPR003610; -.
DR Pfam; PF00704; Glyco_hydro_18; 2.
DR SMART; SM00495; ChtBD3; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 831 AA; 90021 MW; 6982E0B45B72AD31 CRC64;

Query Match          51.1%; Score 45; DB 2; Length 831;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQE 15
   || || || || |
Db 400 WDDSAKAPLYLNPE 413

RESULT 5
Q9I4U2
ID Q9I4U2 PRELIMINARY; PRT; 847 AA.
AC Q9I4U2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PROBABLE PENICILLIN AMIDASE.
GN PA1032.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004535; AAG04421.1; -.
DR InterPro; IPR002692; -.
DR Pfam; PF01804; Penicillamidase; 4.
SQ SEQUENCE 847 AA; 94069 MW; 141B2C463E3071C6 CRC64;

Query Match          51.1%; Score 45; DB 2; Length 847;
Best Local Similarity 52.9%; Pred. No. 26;
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;

QY 1 DW--DTSALAPYLGTQE 15
   :| | :| || | |
Db 721 EQQSDSSKMAPYLGAGE 737

RESULT 6
Q9F9R1
ID Q9F9R1 PRELIMINARY; PRT; 207 AA.
AC Q9F9R1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SUPEROXIDE DISMUTASE.
GN SOD.
OS Mycobacterium paratuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=K-10;
RA Liu X., Feng Z., Cirillo J., Barletta R.G.;
RT "Mycobacterium paratuberculosis manganese superoxide dismutase.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AF180816; AAG09425.1; -.
SQ SEQUENCE 207 AA; 23030 MW; EDA8C2EB40ED428D CRC64;

Query Match          48.9%; Score 43; DB 2; Length 207;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQ 14
   ||| :|| |:: |
Db 10 DWDYAALPHISGQ 23

RESULT 7
Q9I567
ID Q9I567 PRELIMINARY; PRT; 275 AA.
AC Q9I567;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN PA0878.
GN PA0878.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004522; AAG04267.1; -.
KW Hypothetical protein.
SQ SEQUENCE 275 AA; 30521 MW; 5DC323738335DFE CRC64;

Query Match          48.9%; Score 43; DB 2; Length 275;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 SALAPYLGTQEE 16
   || ||::| |||
Db 4 SAFAPWIGRQEE 15

RESULT 8
Q55118
ID Q55118 PRELIMINARY; PRT; 402 AA.
AC Q55118;
DT 01-NOV-1996 (TrEMBLrel. 01, Created).
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 43.9 KDA PROTEIN.
GN SLL0408.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA Tabata S.;
```

RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugiyura M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64% to 92% of the genome.";  
 RN DNA Res. 2:153-166(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hiroseawa M., Sugiyura M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RN DNA Res. 3:109-136(1996).  
 DR EMBL: D64001; BAA10250.1; -  
 DR HSSP: P20752; ICLH  
 DR InterPro; IPR002130; -  
 DR PROSITE; PS50072; CSA\_PPIASE\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 402 AA; 43909 MW; 08A1A3D8A3E9CFAC CRC64;

Query Match 48.9%; Score 43; DB 2; Length 402;  
 Best Local Similarity 53.8%; Pred. No. 26;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 WDTSALAPYLGTQ 14  
 I: I: I: I: I: I:  
 Db 6 WENFAVAPYAGEQ 18

RESULT 9  
 ID Q9X7Q9 PRELIMINARY; PRT; 459 AA.  
 AC Q9X7Q9;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE PUTATIVE SERINE/THREONINE PROTEIN KINASE (FRAGMENT).  
 GN SC5H1.01.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN {1}  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Oliver K., Harris D.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).

DR EMBL: AL049863; CAB42926.1; -  
 DR InterPro; IPR000719; -  
 DR Pfam: PF00069; pkinase; 1.  
 DR PROSITE; PS50011; STYKC; 1.  
 DR SMART; SM00221; STYKC; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 FT NON\_TER 1  
 FT NON\_TER 459  
 SQ SEQUENCE 459 AA; 47539 MW; A7628F4595FA2535 CRC64;

Query Match 48.9%; Score 43; DB 2; Length 459;  
 Best Local Similarity 58.8%; Pred. No. 30;  
 Matches 10; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

OY 1 DWDTs---ALAPYLGT 13  
 I: I: I: I: I: I:  
 Db 437 DWDDSANERAAFLGT 453

RESULT 10  
 ID P79022 PRELIMINARY; PRT; 227 AA.  
 AC P79022;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE SUPEROXIDE DISMUTASE (EC 1.15.1.1) PRECURSOR.  
 GN MNSOD.  
 OS Candida sp.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=39398;  
 RN {1}  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HN95;  
 RA Choi S.-Y., Nam Y.-S., Hong Y.M.;  
 RT "Molecular cloning and characterization of Mn-superoxide dismutase  
 RT gene from Candida sp.";  
 RL Korean J. Microbiol. 35:309-314(1997).  
 CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
 CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
 CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).  
 CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
 CC FAMILY.

DR EMBL: Y11598; CAA72335.1; -  
 DR HSSP: P09214; IMNG.  
 DR InterPro; IPR001189; -  
 DR Pfam: PF00081; sodfe; 1.  
 DR ProDom; PD000475; -; 1.  
 DR PROSITE; PS00088; SOD\_MN; 1.  
 KW Oxidoreductase; Signal.  
 FT SIGNAL 1  
 FT CHAIN 31 227  
 SQ SEQUENCE 227 AA; 25227 MW; 29E7F6E3C7743DB0 CRC64;  
 POTENTIAL.  
 SUPEROXIDE DISMUTASE.

Query Match 47.7%; Score 42; DB 3; Length 227;  
 Best Local Similarity 50.0%; Pred. No. 20;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 DWDTsALAPYLGTQ 14  
 I: I: I: I: I: I:  
 Db 36 DWDFGALEPHISGQ 49

RESULT 11  
 ID Q9PLR4 PRELIMINARY; PRT; 490 AA.  
 AC Q9PLR4;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)



OS Streptomyces spectabilis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=68270;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=NRRL 2494;  
 RA Altenbuchner J., Jennen D., Volff J.N.;  
 RT "The spectinomycin resistance gene of the spectinomycin producer  
 Streptomyces spectabilis";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF170704; AAD50455.1; -  
 SQ SEQUENCE 332 AA; 36128 MW; 76703C681FF9A39A CRC64;

Query Match 46.6%; Score 41; DB 2; Length 332;  
 Best Local Similarity 77.8%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

OY 1 DWDTSALAP 9  
 |||||  
 Db 243 DWDTVGLAP 251

RESULT 16  
 O85977 PRELIMINARY; PRT; 343 AA.  
 ID O85977;  
 AC 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE 4-HYDROXY-2-OXOVALERATE ALDOLASE.  
 GN XYLK.  
 OS Sphingomonas aromaticivorans.  
 OG Plasmid pNL1.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;  
 OC Sphingomonas  
 OX NCBI\_TaxID=48935;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F199;  
 RA Romine M.F., Stillwell L.C., Wong K.-K., Thurston S.J., Sisk E.C.,  
 RS Sersen C.W., Gaasterland T., Saffer J.D., Fredrickson J.K.;  
 RT "Complete sequence of a 184 kb catabolic plasmid from Sphingomonas  
 aromaticivorans strain F199";  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF079317; AAD03993.1; -  
 DR InterPro; IPR000891; -  
 DR Pfam; PF00682; HMGL-like; 1.  
 KW Plasmid.  
 SQ SEQUENCE 343 AA; 36859 MW; 465022794BBCCDF9 CRC64;

Query Match 46.6%; Score 41; DB 2; Length 343;  
 Best Local Similarity 33.3%; Pred. No. 47;  
 Matches 10; Conservative 2; Mismatches 4; Indels 14; Gaps 1;

OY 1 DWD-----TSALAPYLGTQEE 16  
 |||||  
 Db 70 DWDWIEAAADVKNVAVLTLLVPGIGTAAE 99

RESULT 17  
 Q92306 PRELIMINARY; PRT; 343 AA.  
 ID Q92306;  
 AC 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE 4-HYDROXY-2-OXOVALERATE ALDOLASE (4-HYDROXY-2-KETOVALERATE ALDOLASE).  
 GN PHNJ.  
 OS Sphingomonas chungbukensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;

OC Sphingomonas.  
 OX NCBI\_TaxID=56193;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=DJ77;  
 RX MEDLINE=99182313; PubMed=10080921;  
 RA Hwang S., Kim S.J., Kim C.-K., Kim Y., Kim S.-J., Kim Y.-C.;  
 RT "The phnJ genes encoding acetaldehyde dehydrogenase (acylating) and  
 4-hydroxy-2-oxovalerate aldolase in Pseudomonas sp. DJ77 and their  
 evolutionary implications";  
 RL Biochem. Biophys. Res. Commun. 256:469-473(1999).  
 RN [2]  
 RP SEQUENCE OF 294-343 FROM N.A.  
 RC STRAIN=DJ77;  
 RA Kim S., Hwang S.-Y., Lee K.-S., Kim Y., Kim Y.-C.;  
 RT "Nucleotide sequence and phylogenetic analysis of phnKL genes encoding  
 4-oxalocrotonate decarboxylase and 4-oxalocrotonate isomerase in  
 Pseudomonas sp. DJ77";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF061803; AAC95295.1; -  
 DR EMBL; AF073359; AAD03834.1; -  
 DR InterPro; IPR000891; -  
 DR Pfam; PF00682; HMGL-like; 1.  
 SQ SEQUENCE 343 AA; 36863 MW; A6A02A7493D53825 CRC64;

Query Match 46.6%; Score 41; DB 2; Length 343;  
 Best Local Similarity 33.3%; Pred. No. 47;  
 Matches 10; Conservative 2; Mismatches 4; Indels 14; Gaps 1;

OY 1 DWD-----TSALAPYLGTQEE 16  
 |||||  
 Db 70 DWDWIEAAADVKNVAVLTLLVPGIGTAAE 99

RESULT 18  
 Q28450 PRELIMINARY; PRT; 359 AA.  
 ID Q28450;  
 AC 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE INTERPHOTORECEPTOR RETINOID BINDING PROTEIN (FRAGMENT).  
 GN IRBP.  
 OS Loxodonta africana (African elephant).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Proboscidea; Elephantidae; Loxodonta.  
 OX NCBI\_TaxID=9785;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96304325; PubMed=8660440;  
 RA Stanhope M.J., Smith M.R., Waddell V.G., Porter C.A., Shivji M.S.,  
 RA Goodman M.;  
 RT "Mammalian evolution and the interphotoreceptor retinoid binding  
 protein (IRBP) gene: convincing evidence for several superordinal  
 clades";  
 RL J. Mol. Evol. 43:83-92(1996).  
 DR EMBL; U48711; AAB18658.1; -  
 DR InterPro; IPR003581; -  
 DR SMART; SM00245; TSPC; 1.  
 FT NON\_TER  
 FT NON\_TER 359  
 SQ SEQUENCE 359 AA; 39216 MW; 2C759A4621C0D312 CRC64;

Query Match 46.6%; Score 41; DB 6; Length 359;  
 Best Local Similarity 46.7%; Pred. No. 50;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 WDTSALAPYLGTQEE 16  
 |||||  
 Db 201 WEGSGVLPCVGTPEE 215



```
RESULT 19
Q28556          PRELIMINARY;          PRT; 410 AA.
AC Q28556;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE INTERPHOTORECEPTOR RETINOID BINDING PROTEIN (FRAGMENT).
GN IRBP.
OS Orycteropus afer (Aardvark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Tubulidentata; Orycteropodidae; Orycteropus.
OX NCBI_TaxID=9818;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96304325; PubMed=8660440;
RA Stanhope M.J., Smith M.R., Waddell V.G., Porter C.A., Shivji M.S.,
RA Goodman M.;
RT "Mammalian evolution and the interphotoreceptor retinoid binding
RT protein (IRBP) gene: convincing evidence for several superordinal
RT clades.";
RL J. Mol. Evol. 43:83-92(1996).
DR EMBL; U48712; AAB18659.1; -.
DR InterPro; IPR003581; -.
DR SMART; SM00245; TSPC; 1.
FT NON_TER 1
FT NON_TER 410 410
SQ SEQUENCE 410 AA; 44964 MW; 791C6D9F1CE31FAA CRC64;

Query Match 46.6%; Score 41; DB 6; Length 410;
Best Local Similarity 46.7%; Pred. No. 58;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16
   | | | | | | | | | |
Db 252 WEGSGVLPVGTPEE 266

RESULT 20
Q9Y9T8          PRELIMINARY;          PRT; 427 AA.
AC Q9Y9T8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE 427AA LONG HYPOTHETICAL L-ASPARAGINASE.
GN APE2200.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudooh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000063; BAA81212.1; -.
DR InterPro; IPR000267; -.
DR Pfam; PF00710; Asparaginase; 1.
DR PRINTS; PR00139; ASGNLASE.
DR PROSITE; PS00917; ASN_GLN_ASE_2; 1.
SQ SEQUENCE 427 AA; 46373 MW; B932AD12884ED923 CRC64;
```

```
Query Match 46.6%; Score 41; DB 1; Length 427;
Best Local Similarity 46.7%; Pred. No. 61;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQE 15
   | | | | | | | | | |
Db 90 DYETGAVKPYLDASE 104

RESULT 21
Q9N1T0          PRELIMINARY;          PRT; 2487 AA.
AC Q9N1T0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MANNOSE 6-PHOSPHATE/INSULIN-LIKE GROWTH FACTOR 2 RECEPTOR.
GN M6P/IGF2R.
OS Ornithorhynchus anatinus (Duckbill platypus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
OX NCBI_TaxID=9258;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TASMANIAN;
RX MEDLINE=20337954; PubMed=10882106;
RA Killian J.K., Byrd J.C., Jirtle J.V., Munday B.L., Stoskopf M.K.,
RA MacDonald R.G., Jirtle R.L.;
RT "M6P/IGF2R Imprinting Evolution in Mammals.";
RL Mol. Cell 5:707-716(2000).
DR EMBL; AF151172; AAF68173.1; -.
DR InterPro; IPR000479; -.
DR InterPro; IPR000562; -.
DR InterPro; IPR001150; -.
DR InterPro; IPR001596; -.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00878; CIMR_repeat; 12.
DR PRINTS; PR00013; FNTYPEII.
DR PROSITE; PS00023; FIBRONECTIN_2; UNKNOWN_1.
DR PROSITE; PS00850; GLY_RADICAL; UNKNOWN_1.
DR PROSITE; PS00387; PPASE; UNKNOWN_1.
DR SMART; SM00059; FN2; 1.
KW Receptor.
SQ SEQUENCE 2487 AA; 273689 MW; 6CEAEFA9FC32510F CRC64;

Query Match 46.6%; Score 41; DB 6; Length 2487;
Best Local Similarity 46.7%; Pred. No. 4.3e+02;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16
   | | | | | | | | | |
Db 905 WDTEAACPVATTKDE 919

RESULT 22
Q9L4X3          PRELIMINARY;          PRT; 9477 AA.
AC Q9L4X3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE NYSI.
GN NYSI.
OS Streptomyces noursei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
```

RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,  
RA Valla S., Zotchev S.B.;  
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in  
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and  
RT deduction of the biosynthetic pathway.";  
RL Chem. Biol. 7:395-403(2000).  
CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)  
CC FAMILY  
DR EMBL: AF263912; AAF71766.1; -.  
DR InterPro: IPR000255; -.  
DR InterPro: IPR000794; -.  
DR InterPro: IPR001005; -.  
DR InterPro: IPR001227; -.  
DR InterPro: IPR001899; -.  
DR InterPro: IPR002188; -.  
DR Pfam: PF00106; adh\_short; 3.  
DR Pfam: PF00109; ketoacyl-synt; 6.  
DR Pfam: PF00550; pp-binding; 6.  
DR Pfam: PF00698; Acyl-transf; 6.  
DR PROSITE: PS50075; ACP\_DOMAIN; 6.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 6.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 6.  
DR Oxidoreductase; Phosphopantetheine; Transferase.  
KW SEQUENCE 9477 AA; 988151 MW; OD2BCA5D7B265483 CRC64;  
  
Query Match 46.6%; Score 41; DB 2; Length 9477;  
Best Local Similarity 56.2%; Pred NO 1.9e+03; Mismatches 6; Indels 0; Gaps 0;  
Matches 9; Conservative 1;  
  
QY 1 DWDTSALAPYLGTQEE 16  
Db 4267 DGDVSAALTAALGTDED 4282  
  
RESULT 23  
ID O16266 PRELIMINARY; PRT; 258 AA.  
AC O16266;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE F40A3.6 PROTEIN.  
GN F40A3.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smalldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Geisel C., Bradshaw H., Keppler D.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [13]

RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF016423; AAB65325.1; -. 7D6579D97B2E1611 CRC64;  
SQ SEQUENCE 258 AA; 30243 MW; 30243 MW;  
  
Query Match 45.5%; Score 40; DB 5; Length 258;  
Best Local Similarity 63.6%; Pred NO 51; Mismatches 3; Indels 0; Gaps 0;  
Matches 7; Conservative 3;  
  
QY 1 DWDTSALAPYL 11  
Db 201 DFDSSPLSPYL 211  
  
RESULT 24  
ID Q911R6 PRELIMINARY; PRT; 294 AA.  
AC Q911R6;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL PROTEIN PA2201.  
GN PA2201.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL: AE004646; AAG05589.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 294 AA; 32850 MW; EA6F273162A14996 CRC64;  
  
Query Match 45.5%; Score 40; DB 2; Length 294;  
Best Local Similarity 66.7%; Pred. NO. 59; Mismatches 1; Indels 0; Gaps 0;  
Matches 8; Conservative 1;  
  
QY 3 DTSALAPYLGTQ 14  
Db 218 DAQALAPYLQSQ 229  
  
RESULT 25  
ID Q9M6J3 PRELIMINARY; PRT; 392 AA.  
AC Q9M6J3;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE PHYTOCHROME C (FRAGMENT).  
GN PHYC.  
OS Aquilegia sp. SM-1999.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Ranunculales; Ranunculaceae; Aquilegia.  
OX NCBI\_TaxID=112038;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mathews S., Donoghue M.J.;

```
RT "Root of the angiosperm phylogenetic tree inferred from duplicated
RT phytochrome genes."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190067; AAF26313.1; -.
DR InterPro; IPR001294; -.
DR InterPro; IPR003018; -.
DR Pfam; PF00360; phytochrome; 1.
DR Pfam; PF01590; GAF; 1.
DR PRINTS; PRO1033; PHYTOCHROME.
DR PROSITE; PS00245; PHYTOCHROME_1; 1.
DR PROSITE; PS50046; PHYTOCHROME_2; 1.
DR SMART; SM00065; GAF; 1.
FT NON_TER 1
FT NON_TER 392
SQ SEQUENCE 392 AA; 43847 MW; EBF4A728C1EEB529 CRC64;

Query Match 45.5%; Score 40; DB 10; Length 392;
Best Local Similarity 58.3%; Pred. No. 82;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLG 12
Db 75 EWKISDLEPYLG 86

RESULT 26
Q9LHI4 PRELIMINARY; PRT; 440 AA.
AC Q9LHI4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SERINE CARBOXYPEPTIDASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002047; BAB03131.1; -.
DR InterPro; IPR001563; -.
DR Pfam; PF00450; serine_carbpept; 2.
DR PRINTS; PR00724; CRBOXYPTASEC.
KW Carboxypeptidase.
SQ SEQUENCE 440 AA; 49890 MW; D70C6F5E26ED8C3C CRC64;

Query Match 45.5%; Score 40; DB 10; Length 440;
Best Local Similarity 53.8%; Pred. No. 93;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTOE 15
Db 354 DHDMLETPYVGTD 366

RESULT 27
Q59326
ID Q59326 PRELIMINARY; PRT; 482 AA.
```

```
AC Q59326;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CHITINASE PRECURSOR (EC 3.2.1.14) (CHITODEXTRINASE) (1,4-BETA-POLY-N-
DE ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).
GN CHIA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1237;
RA Fuchs K.P., Schwarz W.H., Staudenbauer W.L.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-
CC ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; Z68924; CAA93150.1; -.
DR HSSP; P07254; ICTN.
DR InterPro; IPR001223; -.
DR InterPro; IPR001579; -.
DR InterPro; IPR002048; -.
DR InterPro; IPR002105; -.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF00704; Glyco_hydro_18; 2.
DR PROSITE; PS01095; CHITINASE_18; 1.
DR PROSITE; PS00448; CLOS_CELLULOsome_RPT; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 482
SQ SEQUENCE 482 AA; 55028 MW; BF047D8DE3588DF5 CRC64;

Query Match 45.5%; Score 40; DB 2; Length 482;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 WDTSAAPYLGTOE 15
Db 341 WDEYAMVPYLWNPE 354

RESULT 28
Q9FFL1 PRELIMINARY; PRT; 490 AA.
AC Q9FFL1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SIMILARITY TO CATIONIC AMINO ACID TRANSPORTER 1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT p1 clones.";
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005237; BAB09657.1; -.
SQ SEQUENCE 490 AA; 53693 MW; E478743784CFE309 CRC64;
```

Query Match 45.5%; Score 40; DB 10; Length 490;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQE 15  
Db 120 WVTLAGMPYWGFOQ 133

RESULT 29  
Q9XE45 PRELIMINARY; PRT; 732 AA.  
AC Q9XE45;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PUTATIVE REVERSE TRANSCRIPTASE.  
GN TSM2.3.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,  
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
RA Fraser C.M., Venter J.C.;  
RA "Arabidopsis thaliana Chromosome II BAC T5M2 genomic sequence.";  
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE  
TRANSCRIPTASE).  
CC EMBL; AC007730; AAD37021.1;  
DR InterPro; IPR000477;  
DR Pfam; PF00078; rvt; 2.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 732 AA; 83450 MW; 3D58BF00D56E8F2E CRC64;

Query Match 45.5%; Score 40; DB 10; Length 732;  
Best Local Similarity 41.7%; Pred. No. 1.7e+02;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGT 13  
Db 500 WDMATIAPYISS 511

RESULT 30  
Q9SDA6 PRELIMINARY; PRT; 957 AA.  
AC Q9SDA6;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PUTATIVE DISEASE RESISTANCE PROTEIN.  
GN AT2G17050.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RA "Sequence and analysis of chromosome II of Arabidopsis thaliana.";  
RL Nature 402:761-768(1999).  
DR EMBL; AC002354; AAF18599.1;  
DR InterPro; IPR000157;  
DR InterPro; IPR000767;  
DR InterPro; IPR001611;  
DR InterPro; IPR002182;  
DR Pfam; PF00560; LRR; 7;  
DR Pfam; PF00931; NB-ARC; 1.  
DR Pfam; PF01582; TIR; 1.  
DR PRINTS; PR00364; DISEASERSIST.  
DR SMART; SM00255; TIR; 1.  
SQ SEQUENCE 957 AA; 109859 MW; 5A8E1B33F3B97E34 CRC64;

Query Match 45.5%; Score 40; DB 10; Length 957;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQE 15  
Db 335 WDTSIQPLEDEE 348

RESULT 31  
Q9VSC9 PRELIMINARY; PRT; 1069 AA.  
AC Q9VSC9;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CG7565 PROTEIN.  
GN CG7565.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA MEDLINE=20196006; PubMed=107311132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balslev D.,  
RA Balleson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,





RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weltri R., Goo Y.A.,  
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Madden D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.,  
RT "Genome sequence of Halobacterium species NRC-1";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE004988; AAG18841.1; --  
DR InterPro; IPR001656; --  
DR Pfam; PF01142; UPF0024; 1.  
SQ SEQUENCE 575 AA; 63009 MW; C0172D520F687808 CRC64;

Query Match 44.9%; Score 39.5; DB 1; Length 575;  
Best Local Similarity 50.0%; Pred. No. 1.5e+02; Indels 1; Gaps 1;  
Matches 8; Conservative 3; Mismatches 4;

OY 1 DWDTLSALAPYLGTQEE 16  
Db 338 DWEAANA-YLGAPTE 352

RESULT 35  
O53813 PRELIMINARY; PRT; 81 AA.  
AC O53813;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE HYPOTHETICAL 8.7 KDA PROTEIN.  
GN RV0750 OR MV041.24.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence";  
RL Nature 393:537-544(1998).  
DR EMBL; AL021958; CAAL7517.1; --  
DR TuberculList; RV0750; --  
KW Hypothetical protein.  
SQ SEQUENCE 81 AA; 8719 MW; A02EDE63B68DAFEA CRC64;

Query Match 44.3%; Score 39; DB 2; Length 81;  
Best Local Similarity 70.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 WDTLSALAPYL 11  
Db 60 WYTOGLAPYL 69

RESULT 36  
Q9LGN9 PRELIMINARY; PRT; 84 AA.  
ID Q9LGN9  
AC Q9LGN9;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE P0684C01.6 PROTEIN.  
GN P0684C01.6.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;  
OC Oryza.  
OX NCHI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RC Sasaki T., Matsumoto T., Yamamoto K.;  
RA "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 1, PAC clone:  
RT P0684C01";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AP002487; BAB07945.1; --  
SQ SEQUENCE 84 AA; 9085 MW; 33CA61DF0A6CCE0C CRC64;

Query Match 44.3%; Score 39; DB 10; Length 84;  
Best Local Similarity 77.8%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 DWDTLSALAP 9  
Db 62 DWDTSMEEP 70

RESULT 37  
O53803 PRELIMINARY; PRT; 175 AA.  
AC O53803;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE HYPOTHETICAL 19.5 KDA PROTEIN.  
GN RV0740 OR MV041.14.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence";  
RL Nature 393:537-544(1998).  
DR EMBL; AL021958; CAAL7507.1; --  
DR TuberculList; RV0740; --  
KW Hypothetical protein.  
SQ SEQUENCE 175 AA; 19452 MW; CD8EB84AE02A6723 CRC64;

Query Match 44.3%; Score 39; DB 2; Length 175;  
Best Local Similarity 70.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 WDTLSALAPYL 11  
Db 154 WYTOGLAPYL 163

RESULT 38  
P78203

ID P78203 PRELIMINARY; PRT; 179 AA.  
AC P78203;  
DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE BENZENE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.3) (FRAGMENT).  
GN BEDC1.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97349980; PubMed=9205837;  
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features.";  
RL DNA Res. 4:91-113(1997).  
DR EMBL; D90883; BAA16433.1; -.  
DR HSSP; P23094; 1NDO.  
DR InterPro; IPR001281; -.  
DR InterPro; IPR001663; -.  
DR Pfam; PF00355; Rieske; 1.  
DR PRINTS; PR00090; RINGDIOXGNASE.  
DR PROSITE; PS00570; RING\_HYDROXYL\_ALPHA; 1.  
FT NON\_TER 179  
SQ SEQUENCE 179 AA; 20454 MW; 78FDA1562601573D CRC64;

Query Match 44.3%; Score 39; DB 2; Length 179;  
Best Local Similarity 64.3%; Pred. No. 51;  
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

OY 1 DWDTSA--LAPYLG 12  
:||||| | |||  
Db 155 NWDTSAPGLRDYLG 168

RESULT 39  
P79827  
ID P79827 PRELIMINARY; PRT; 196 AA.  
AC P79827;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE CATHEPSIN L (EC 3.4.22.15) (FRAGMENT).  
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
OX NCBI\_TaxID=8022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RA Nickel X.F., An H., Li F., Seymour T.A., Morrissey M.T., Barnes D.W.;  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: SPECIFICITY CLOSE TO THAT OF PAPAIN.  
DR EMBL; U61296; AAB37252.1; -.  
DR HSSP; P07711; 1CJL.  
DR MEROPS; C01.032; -.  
DR InterPro; IPR000169; -.  
DR InterPro; IPR000668; -.  
DR Pfam; PF00112; Peptidase\_C1; 1.  
DR PRINTS; PR00705; PAPAIN.  
DR PROSITE; PS00640; THIOLESTEROIDASE; 1.  
DR PROSITE; PS00639; THIOLESTEROIDASE; 1.

KW Hydrolase; Thiol protease.  
FT NON\_TER 1  
SQ SEQUENCE 196 AA; 21537 MW; 2AB15321E1E4779E CRC64;

Query Match 44.3%; Score 39; DB 13; Length 196;  
Best Local Similarity 50.0%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 DTSALAPYLGTQEE 16  
|| ||: || |:  
Db 58 DTEESYPYVGTD 71

RESULT 40  
Q9L0K0  
ID Q9L0K0 PRELIMINARY; PRT; 209 AA.  
AC Q9L0K0;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE PUTATIVE SECRETED PROTEIN.  
GN SCD40A.11C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL161691; CAB81856.1; -.  
SQ SEQUENCE 209 AA; 21951 MW; 63BB4EEFFB9718BC CRC64;

Query Match 44.3%; Score 39; DB 2; Length 209;  
Best Local Similarity 66.7%; Pred. No. 61;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 WDTALAPY 10  
| | | | |  
Db 121 WSTAAPY 129

RESULT 41  
Q9GJT2  
ID Q9GJT2 PRELIMINARY; PRT; 282 AA.  
AC Q9GJT2;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE ESTERASE D.  
GN PIGESD OR ESD.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-OHMINI;  
RA Omi T., Tsuchida S., Kajli E.;  
RT "Sus scrofa cDNA for Esterase D (phenotype A), complete cds."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Tsuchida S., Omi T., Ikemoto S.;  
RT "Cloning of swine esterase D cDNA."  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038362; BAB11922.1; -  
DR EMBL; AF254785; AAG17630.1; -  
SQ SEQUENCE 282 AA; 31482 MW; 62846B9291591FBA CRC64;

Query Match 44.3%; Score 39; DB 6; Length 282;  
Best Local Similarity 50.0%; Pred. No. 85;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQE 15  
Db 183 WGKKAFSGYLGTD 196

RESULT 42  
O54397 PRELIMINARY; PRT; 326 AA.  
AC O54397;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
DE PLASMID PBD2 ORF1.  
OS Rhodococcus erythropolis.  
OG Plasmid PBD2.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.  
OX NCBI\_TaxID=1833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BD2;  
RA Kessler M., Averhoff B., Gottschalk G.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U83846; AAB94318.1; -  
KW Plasmid.  
SQ SEQUENCE 326 AA; 34702 MW; 9CD534D5B2A19298 CRC64;

Query Match 44.3%; Score 39; DB 2; Length 326;  
Best Local Similarity 77.8%; Pred. No. 1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 TSALAPYL 12  
Db 190 TSGLAPYIG 198

RESULT 43  
O16739 PRELIMINARY; PRT; 337 AA.  
AC O16739;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE F45C12.8 PROTEIN.  
GN F45C12.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Johnston L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P., III of C.  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans."  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Johnson D., Kramer J., Keppler D.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF016684; AAB66207.1; -  
DR InterPro; IPR001810; -  
DR InterPro; IPR002900; -  
DR Pfam; PF00646; F-box; 1;  
DR Pfam; PF01827; DUF38; 2;  
DR PROSITE; PS50181; FBOX; 1;  
DR SMART; SM00256; FBOX; 1;  
SQ SEQUENCE 337 AA; 39534 MW; FB45E9DEAB390DB4 CRC64;

Query Match 44.3%; Score 39; DB 5; Length 337;  
Best Local Similarity 50.0%; Pred. No. 1e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 DWDTALAPYLGTQEE 16  
Db 286 DMNTSLKSVFGTQDE 301

RESULT 44  
O58813 PRELIMINARY; PRT; 346 AA.  
AC O58813;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE 346AA LONG HYPOTHETICAL ASPARTOKINASE.  
GN PH1086.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,  
RA Acki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."  
RL DNA Res. 5:55-76(1998).  
RL EMBL; AP000004; BAA30185.1; -  
DR InterPro; IPR001048; -  
DR InterPro; IPR001341; -  
DR Pfam; PF00696; aakinas; 1.  
DR PROSITE; PS00324; ASPARTOKINASE; UNKNOWN\_1.  
SQ SEQUENCE 346 AA; 38543 MW; F5958F0ECFA7470D CRC64;



Query Match 44.3%; Score 39; DB 1; Length 346;  
Best Local Similarity 77.8%; Pred. No. 1.1e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 DTSALAPYL 11  
|||:|:|  
Db 80 DTSSLSPYL 88

RESULT 45  
Q9HP38  
ID Q9HP38 PRELIMINARY; PRT; 347 AA.  
AC Q9HP38;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE ALCOHOL DEHYDROGENASE.  
GN ADH4 OR VNG1821G.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
OX NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
RT "Genome sequence of Halobacterium species NRC-1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE005083; AAG20032.1; -.  
DR InterPro; IPR002085; -.  
DR Pfam; PF00107; adh\_zinc; 1.  
SQ SEQUENCE 347 AA; 36719 MW; 5352453D854E26E1 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 347;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DWDTSALAP 9  
||:|:|  
Db 140 DWETAAAP 148

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:54:12 ; Search time 66.55 Seconds  
(without alignments)  
10.931 Million cell updates/sec

Title: US-09-439-313-558  
Perfect score: 68  
Sequence: 1 AGLLCPDPRPLE 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	100.0	553	19	AAW71869 Amino acid encoded
2	68	100.0	553	19	AAW69385 Prostate tumour sp
3	68	100.0	553	21	AAW28527 Protein encoded by
4	68	100.0	553	21	AAW82002 Human immunogenic
5	45	66.2	191	21	AAW42089 Human ORFX ORF1853
6	41	60.3	15	21	AAW98874 HLA class II bindi
7	41	60.3	15	21	AAW98970 HLA class II bindi
8	41	60.3	59	21	AAW21202 Human HER-2/neu pr
9	41	60.3	266	21	AAW21201 Human HER-2/neu pr
10	41	60.3	291	19	AAW63682 Human secreted pro
11	41	60.3	293	21	AAW58875 Breast and ovarian

12	41	60.3	568	16	AAW65496 Marek's disease vi
13	41	60.3	712	21	AAW21204 Human HER-2/neu fu
14	41	60.3	729	22	AAU00020 Human Plexin prote
15	41	60.3	919	21	AAW21203 Human HER-2/neu fu
16	41	60.3	1200	21	AAW21208 Human HER-2/neu pr
17	41	60.3	1255	17	AAW01111 HER-2/neu protein.
18	41	60.3	1255	20	AAW92406 Human HER-2/neu on
19	41	60.3	1255	21	AAW21198 Human HER-2/neu pr
20	41	60.3	1255	21	AAW84780 Amino acid sequenc
21	41	60.3	1255	21	AAW92620 Human heregulin 2
22	41	60.3	1255	22	AAW60167 HER2 transgene pla
23	41	60.3	1433	14	AAW39568 Sequence of c-erbB
24	41	60.3	2135	22	AAU00019 Human Plexin prote
25	40	58.8	122	12	AAW15105 hCG/BLH chimera, D
26	40	58.8	328	17	AAW05504 HCMV Toledo strain
27	40	58.8	485	21	AAW31001 Arabidopsis thalia
28	40	58.8	486	21	AAW31000 Arabidopsis thalia
29	40	58.8	522	20	AAW38566 Neisseria meningit
30	40	58.8	522	20	AAW38568 Neisseria gonorrh
31	40	58.8	523	20	AAW38567 Neisseria meningit
32	40	58.8	526	21	AAW74891 Neisseria meningit
33	40	58.8	533	21	AAW74890 Neisseria gonorrh
34	40	58.8	533	21	AAW74892 Neisseria meningit
35	40	58.8	538	21	AAW30999 Arabidopsis thalia
36	39	57.4	161	21	AAW22586 Zea mays protein f
37	39	57.4	172	21	AAW33891 Arabidopsis thalia
38	39	57.4	198	20	AAW18084 Histamine binding
39	39	57.4	264	21	AAW27731 Sequence homologou
40	39	57.4	285	19	AAW77301 Amino acid sequenc
41	39	57.4	379	21	AAW59390 Murine soluble int
42	39	57.4	392	21	AAW42988 Human ORFX ORF2752
43	39	57.4	432	17	AAW92813 Murine interleukin
44	39	57.4	432	22	AAW36653 Mouse IL-11 recept
45	39	57.4	441	17	AAW99091 Murine Etl-2 gene
46	39	57.4	454	21	AAW25433 Arabidopsis thalia
47	39	57.4	454	21	AAW34331 Zea mays protein f
48	39	57.4	487	21	AAW17934 Arabidopsis thalia
49	39	57.4	495	21	AAW34330 Zea mays protein f
50	39	57.4	500	21	AAW17933 Arabidopsis thalia
51	39	57.4	503	21	AAW25432 Arabidopsis thalia
52	39	57.4	521	20	AAW02367 Polypeptide identi
53	39	57.4	527	21	AAW34329 Zea mays protein f
54	39	57.4	532	21	AAW17932 Arabidopsis thalia
55	39	57.4	535	21	AAW25431 Arabidopsis thalia
56	39	57.4	733	21	AAW52241 Arabidopsis thalia
57	39	57.4	779	21	AAW52240 Arabidopsis thalia
58	39	57.4	811	21	AAW52239 Arabidopsis thalia
59	39	57.4	1529	21	AAW96744 A. terreus ORF1 es
60	38	55.9	59	20	AAW41511 Fragment of human
61	38	55.9	77	21	AAW02476 Human secreted pro
62	38	55.9	190	17	AAW90768 FGF receptor ligan
63	38	55.9	331	21	AAW41847 Human ORFX ORF1611
64	38	55.9	336	21	AAW27990 Human secreted pro
65	38	55.9	336	22	AAW64536 Human secreted pro

ALIGNMENTS

RESULT	1
AAW71869	
ID	AAW71869 standard; Protein; 553 AA.
XX	
AC	AAW71869;
XX	
DT	06-JAN-1999 (first entry)
XX	
DE	Amino acid encoded by prostate tumour clone L1-12.
XX	
KW	Prostate; cancer; tumour; vaccine; immunogen; clone.
XX	
OS	Homo sapiens.
XX	

PN WO9837093-A2.  
XX  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03492.  
XX  
PR 09-FEB-1998; 98US-0020956.  
XX  
PR 25-FEB-1997; 97US-0806099.  
XX  
PR 01-AUG-1997; 97US-0904804.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
PI WPI; 1998-609886/51.  
XX  
DR N-PSDB; AAV61201.  
XX  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
XX  
PS Example 1; Page 82-84; 130pp; English.  
XX  
CC The present sequence is an immunogenic portion of a prostate tumour  
CC protein. The immunogen, or the DNA encoding it, can be used as a  
CC vaccine for the treatment of prostate cancer. The immunogen was  
CC isolated from a prostate tumour cDNA library obtained by subtracting  
CC a prostate tumour cDNA expression library with a normal tissue cDNA  
CC library.  
XX  
SQ Sequence 553 AA;  
  
Query Match 100.0%; Score 68; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.0076;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AGLLCPDPRPLE 12  
Db |||||  
110 agllcpdprple 121  
  
RESULT 2  
AAW69385  
ID AAW69385 standard; Protein; 553 AA.  
XX  
AC AAW69385;  
XX  
DT 08-DEC-1998 (first entry)  
XX  
DE Prostate tumour specific gene clone L1-12 protein.  
XX  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
XX therapy.  
XX Homo sapiens.  
XX OS  
XX PN WO9837418-A2.  
XX  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03690.  
XX  
PR 09-FEB-1998; 98US-0904809.  
XX  
PR 25-FEB-1997; 97US-0806596.  
XX  
PR 01-AUG-1997; 97US-0904809.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
PI WPI; 1998-480805/41.  
XX  
DR N-PSDB; AAV58586.  
XX

PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
XX  
PS Example 1; Page 87-89; 141pp; English.  
XX  
CC This sequence is encoded by a human prostate tumour specific gene, and  
CC can be used in the method of the invention. The method is for detecting  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC this protein sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
XX  
SQ Sequence 553 AA;  
  
Query Match 100.0%; Score 68; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.0076;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AGLLCPDPRPLE 12  
Db |||||  
110 agllcpdprple 121  
  
RESULT 3  
AAB28527  
ID AAB28527 standard; Protein; 553 AA.  
XX  
AC AAB28527;  
XX  
DT 07-FEB-2001 (first entry)  
XX  
DE Protein encoded by human breast tumour cDNA clone P501S.  
XX  
KW Human; breast tumour antigen; cytostatic; immunotherapy;  
XX breast cancer; vaccine.  
XX Homo sapiens.  
XX OS  
XX PN WO200061756-A2.  
XX  
PD 19-OCT-2000.  
XX  
PF 10-APR-2000; 2000WO-US09688.  
XX  
PR 09-APR-1999; 99US-0288950.  
XX  
PR 02-JUL-1999; 99US-0346327.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Reed SG, Xu J, Dillon DC;  
XX  
PI WPI; 2000-638568/61.  
XX  
DR N-PSDB; AAC79473.  
XX  
PT A novel isolated polypeptide comprising an immunogenic portion of a  
PT breast cancer protein useful in the detection and treatment of breast  
PT cancer -  
XX  
PS Claim 2; Page 92-93; 95pp; English.  
XX  
CC The present sequence is encoded by a cDNA sequence which was isolated  
CC from a breast tumour cDNA library. It is provided in a specification  
CC relating to compounds for immunotherapy and diagnosis of breast cancer.  
CC Breast tumour antigens and the polynucleotides that encode them may be  
CC used in the production of a pharmaceutical composition to be used in the  
CC treatment of breast cancer. Proliferated r cells and incubated antigen  
CC presenting cells are also required. The polypeptides and polynucleotides  
CC may also be used to produce a vaccine.  
XX





CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 191 AA;

Query Match. 66.2%; Score 45; DB 21; Length 191;  
Best Local Similarity 77.8%; Pred. NO. 9.8;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 GLLCPDPRP 10  
| | | | | | | |  
Db 9 gllcpqpqp 17

RESULT 6  
AAAY98874  
ID AAY98874 standard; Peptide; 15 AA.

XX AC AAY98874;

XX DT 07-AUG-2000 (first entry)

XX DE HLA class II binding antigen epitope peptide #63.

XX KW Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;  
KW immune response; chronic viral disease; cancer; autoimmune disease;  
KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;  
KW allograft rejection; allergy; Lyme disease; hepatitis; prostate cancer;  
KW glomerulonephritis; food hypersensitivity; malaria.

XX OS Unidentified.

XX PN WO9961916-A1.

XX PD 02-DEC-1999.

XX PF 28-MAY-1999; 99WO-US12066.

XX PR 29-MAY-1998; 98US-0087192.

XX PA (EPIM-) EPIMMUNE INC.

XX PI Sette A, Southwood S, Sidney J;

XX DR WPI; 2000-097143/08.

XX New compositions containing immunogenic peptide epitopes for various  
XX HLA class II DR molecules useful for inducing helper T cell response

XX Claim 1; Page 41; 60pp; English.

XX The present invention relates to a new pharmaceutical composition  
XX comprising a unit dose form of a peptide, or analogue, comprising an  
XX epitope selected from those represented by peptides AAY98812-Y99339  
XX which are derived from various antigens for various human leucocyte  
XX antigen class DR molecules, representative of the world wide population.  
XX The peptide/analogue binds to an HLA class II molecule at an IC-50 of  
XX less than or equal to 1,000 nM. The pharmaceutical can be used to induce  
XX a helper T cell response. The pharmaceutical focuses the immune response  
XX towards selected determinants and could therefore be used in cases of  
XX chronic viral diseases and cancer. Examples of diseases that can be  
XX treated using the peptide containing pharmaceutical include autoimmune  
XX diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia  
XX gravis), allograft rejection, allergies, Lyme disease, hepatitis,  
XX post-streptococcal endocarditis or glomerulonephritis and food  
XX hypersensitivities. The peptide epitopes can be used to enhance immune  
XX responses against other immunogens administered with the peptides.  
XX Diseases which can be treated using immunogenic mixtures include  
XX prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,  
XX cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may

CC also be used to make monoclonal antibodies useful as potential diagnostic  
CC or therapeutic agents. The peptides may also be useful as diagnostic  
CC reagents, for example, to determine the susceptibility of an individual  
CC to a treatment regimen. Also, the peptides may be used to predict which  
CC individuals will be at substantial risk of developing chronic infection.  
CC The selection of appropriate T and B cell epitopes should allow the  
CC development of epitope based vaccines particularly towards conserved  
CC epitopes of pathogens which are characterized by high sequence  
CC variability such as HIV, HCV and Malaria.

XX Sequence 15 AA;

Query Match. 60.3%; Score 41; DB 21; Length 15;  
Best Local Similarity 66.7%; Pred. NO. 4;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLLCPDPRP 10  
| | | | | | | |  
Db 2 gllcpdpap 10

RESULT 7

AAAY98970

ID AAY98970 standard; Peptide; 15 AA.

XX AC AAY98970;

XX DT 07-AUG-2000 (first entry)

XX DE HLA class II binding antigen epitope peptide #159.

XX KW Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;  
KW immune response; chronic viral disease; cancer; autoimmune disease;  
KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;  
KW allograft rejection; allergy; Lyme disease; hepatitis; prostate cancer;  
KW glomerulonephritis; food hypersensitivity; malaria.

XX OS Unidentified.

XX PN WO9961916-A1.

XX PD 02-DEC-1999.

XX PF 28-MAY-1999; 99WO-US12066.

XX PR 29-MAY-1998; 98US-0087192.

XX PA (EPIM-) EPIMMUNE INC.

XX PI Sette A, Southwood S, Sidney J;

XX DR WPI; 2000-097143/08.

XX New compositions containing immunogenic peptide epitopes for various  
XX HLA class II DR molecules useful for inducing helper T cell response

XX Claim 1; Page 42; 60pp; English.

XX The present invention relates to a new pharmaceutical composition  
XX comprising a unit dose form of a peptide, or analogue, comprising an  
XX epitope selected from those represented by peptides AAY98812-Y99339  
XX which are derived from various antigens for various human leucocyte  
XX antigen class DR molecules, representative of the world wide population.  
XX The peptide/analogue binds to an HLA class II molecule at an IC-50 of  
XX less than or equal to 1,000 nM. The pharmaceutical can be used to induce  
XX a helper T cell response. The pharmaceutical focuses the immune response  
XX towards selected determinants and could therefore be used in cases of  
XX chronic viral diseases and cancer. Examples of diseases that can be  
XX treated using the peptide containing pharmaceutical include autoimmune  
XX diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia  
XX gravis), allograft rejection, allergies, Lyme disease, hepatitis,  
XX post-streptococcal endocarditis or glomerulonephritis and food

CC hypersensitivities. The peptide epitopes can be used to enhance immune  
CC responses against other immunogens administered with the peptides.  
CC Diseases which can be treated using immunogenic mixtures include  
CC prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,  
CC cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may  
CC also be used to make monoclonal antibodies useful as potential diagnostic  
CC or therapeutic agents. The peptides may also be useful as diagnostic  
CC reagents, for example, to determine the susceptibility of an individual  
CC to a treatment regimen. Also, the peptides may be used to predict which  
CC individuals will be at substantial risk of developing chronic infection.  
CC The selection of appropriate T and B cell epitopes should allow the  
CC development of epitope based vaccines particularly towards conserved  
CC epitopes of pathogens which are characterized by high sequence  
CC variability such as HIV, HCV and Malaria.

XX  
SQ Sequence 15 AA;

Query Match 60.3%; Score 41; DB 21; Length 15;  
Best Local Similarity 66.7%; Pred. No. 4;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10  
| | | | |  
Db 2 gffcpdpap 10

RESULT 8  
AAB21202  
ID AAB21202 standard; protein; 59 AA.

XX  
AC AAB21202;

XX  
DT 12-JAN-2001 (first entry)

XX  
DE Human HER-2/neu protein phosphorylation domain partial sequence.

XX  
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.

XX  
OS Homo sapiens.

XX  
PN WO200044899-A1.

XX  
PD 03-AUG-2000.

XX  
PF 28-JAN-2000; 2000WO-US02164.

XX  
PR 29-JAN-1999; 99US-0117976.

XX  
PA (CORI-) CORIXA CORP.

PA (SMIK ) SMITHKLINE BEECHAM.

XX  
PI Cheever MA, Gheysen D;

XX  
DR WPI; 2000-505976/45.

XX  
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -

XX  
PS Claim 27; Fig 11; 128pp; English.

XX  
CC The present sequence is a preferred portion of the phosphorylation domain  
CC of the HER-2/neu protein. HER-2/neu is a member of the tyrosine kinase  
CC family of receptor-like glycoproteins and shows homology to the epidermal  
CC growth factor receptor (EGFR). It probably plays a part in cell growth  
CC and/or differentiation. The HER-2/neu gene is an oncogene. An HER-2/neu  
CC fusion protein comprising a HER-2/neu extracellular domain fused to a  
CC HER-2/neu phosphorylation domain may be used to treat or prevent cancer  
CC by eliciting or enhancing an immune response to the HER-2/neu protein. It  
CC may be used to treat malignancies such as breast, ovarian, colon, lung

CC and prostate cancers, and may be used as an antigen to vaccinate against  
CC these neoplasias.

XX  
SQ Sequence 59 AA;

Query Match 60.3%; Score 41; DB 21; Length 59;  
Best Local Similarity 66.7%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10  
| | | | |  
Db 40 gffcpdpap 48

RESULT 9  
AAB21201  
ID AAB21201 standard; protein; 266 AA.

XX  
AC AAB21201;

XX  
DT 12-JAN-2001 (first entry)

XX  
DE Human HER-2/neu protein phosphorylation domain.

XX  
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.

XX  
OS Homo sapiens.

XX  
PN WO200044899-A1.

XX  
PD 03-AUG-2000.

XX  
PF 28-JAN-2000; 2000WO-US02164.

XX  
PR 29-JAN-1999; 99US-0117976.

XX  
PA (CORI-) CORIXA CORP.

PA (SMIK ) SMITHKLINE BEECHAM.

XX  
PI Cheever MA, Gheysen D;

XX  
DR WPI; 2000-505976/45.

XX  
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -

PS Claim 2; Fig 10; 128pp; English.

XX  
CC The present sequence is the phosphorylation domain of the HER-2/neu  
CC protein. HER-2/neu is a member of the tyrosine kinase family of  
CC receptor-like glycoproteins and shows homology to the epidermal growth  
CC factor receptor (EGFR). It probably plays a part in cell growth and/or  
CC differentiation. The HER-2/neu gene is an oncogene. An HER-2/neu fusion  
CC protein comprising a HER-2/neu extracellular domain fused to a HER-2/neu  
CC phosphorylation domain may be used to treat or prevent cancer by  
CC eliciting or enhancing an immune response to the HER-2/neu protein. It  
CC may be used to treat malignancies such as breast, ovarian, colon, lung  
CC and prostate cancers, and may be used as an antigen to vaccinate against  
CC these neoplasias.

XX  
SQ Sequence 266 AA;

Query Match 60.3%; Score 41; DB 21; Length 266;  
Best Local Similarity 66.7%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10  
| | | | |

Db 40 gffcpdpap 48

RESULT 10  
AAW63682  
ID AAW63682 standard; Protein; 291 AA.  
XX  
AC AAW63682;  
XX  
DT 24-SEP-1998 (first entry)  
XX  
DE Human secreted protein 2.  
XX  
KW Secreted protein; human; cell proliferation; cytokine activity;  
KW tissue growth; cellular differentiation; regeneration; activin;  
KW inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;  
KW anti-inflammatory activity; biomarker.  
XX  
OS Homo sapiens.  
XX  
PN WO9825959-A2.  
XX  
PD 18-JUN-1998.  
XX  
PF 11-DEC-1997; 97WO-US22787.  
XX  
PR 11-DEC-1996; 96US-0032757.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;  
XX  
DR WPI; 1998-348453/30.  
XX  
N-PSDB; AAV43602.  
XX  
Secreted human polypeptides - having cytokine, cell proliferation or  
differentiation, activin or inhibin, tumour inhibition or  
anti-inflammatory activities  
XX  
PS Claim 1; Pages 49-50; 78pp; English.  
XX  
This represents a human secreted protein. The specification provides  
secreted protein sequences (AAW63681 to AAW63699) encoded by the nucleic  
acid sequences shown in AAV43601 to AAV43619. The invention provides a  
method of identifying a secreted polypeptide which is modified by rough  
microsomes. The secreted proteins can be used in assays to determine  
biological activities, such as cytokine, cell proliferation, or cellular  
differentiation activities, tissue growth or regeneration, activin or  
inhibin activity, chemotactic or chemokinetic activity, haemostatic or  
thrombolytic activity, receptor/ligand activity, tumour inhibition, or  
anti-inflammatory activity. The proteins can also be used as  
biomarkers, to identify tissues or cell types which express the proteins,  
or a stage- or disease-specific alteration in protein expression. They  
can be used in protein interaction assays, to identify ligands or binding  
proteins. Compounds which affect the biological activities of the  
secreted proteins or their ability to interact with specific ligands can  
be identified using the proteins in screening assays. The proteins and  
antibodies that bind specifically to the protein can also be used to  
design diagnostic tests and therapeutic compositions for diseases which  
may be associated with altered expression of these proteins. Fusion  
proteins comprising, e.g. signal sequences or transmembrane domains of  
the proteins can be used to target other protein domains to cellular  
membrane or they can be secreted extracellularly.

Sequence 291 AA;

Query Match 60.3%; Score 41; DB 19; Length 291;  
Best Local Similarity 72.7%; Pred. No. 59;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPPDPPL 12  
| | | | |

Db 50 gqlcwsprple 60

RESULT 11  
AAB58875  
ID AAB58875 standard; Protein; 293 AA.  
XX  
AC AAB58875;  
XX  
DT 27-MAR-2001 (first entry)  
XX  
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 583.  
XX  
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
KW neoplastic; neurproliferative; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antiinflammatory; antitumor; antitumor; antitumor;  
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;  
KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
KW cardiovascular disorder; wound healing; neurological disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200055173-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05881.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-611515/58.  
XX  
N-PSDB; AAF21778.  
XX  
New human breast and ovarian cancer associated gene sequences and the  
polypeptides encoded by these genes, useful in the prevention,  
treatment and diagnosis of cancer, immune disorders, cardiovascular  
disorders and neurological diseases -  
XX  
PS Claim 11; Page 1019-1020; 1299pp; English.  
XX  
Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
proteins AAB58711 - AAB59128. The DNA and protein sequences are  
associated with breast and ovarian cancer. Included in the invention are  
sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
isolation and characterisation of the DNA and protein sequences of the  
invention. The breast and ovarian cancer associated DNA, protein, agonist  
or antagonist sequences exhibit cytostatic; immunosuppressive;  
neoplastic; neurproliferative; antiviral; antiallergic; hepatotropic;  
antidiabetic; antiinflammatory; antitumor; antitumor; antitumor;  
KW antibacterial; antifungal; antiparasitic and cardiac activity. The  
polynucleotide and protein sequences are used in the diagnosis of cancer,  
particularly breast and ovarian cancer. The nucleic acid sequences,  
proteins, agonists and antagonists may also be used in the diagnosis,  
prevention and treatment of immune disorders e.g. Addison's disease,  
allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
arthritis and ulcerative colitis; cardiovascular disorders such as  
myocardial ischaemias; wound healing; neurological diseases such as  
cerebral anoxia and epilepsy; and infectious diseases.

Sequence 293 AA;

Query Match 60.3%; Score 41; DB 21; Length 293;  
Best Local Similarity 66.7%; Pred. No. 60;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY      2 GLLCPPRP 10
      | | | | |
Db     113 gffcpbpap 121
```

RESULT 12  
AAR65496  
ID AAR65496 standard; Protein: 568 AA.

RESULT	13
AAB21204	
ID	AAB21204 standard; protein; 712 AA.
XX	
AC	AAB21204;
XX	
DT	12-JAN-2001 (first entry)
XX	
DE	Human HER-2/neu fusion protein.
XX	
KW	Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW	breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW	colon cancer; fusion protein.
XX	
OS	Homo sapiens.

OS Synthetic.  
XX  
PN WO200044899-A1.  
XX  
PD 03-AUG-2000.

RESULT	14	
AAU00020		
ID	AAU00020	standard; Protein; 729 AA.
XX		
AC	AAU00020;	
XX		
DT	09-MAY-2001	(first entry)
XX		
DE	Human Plexin protein.	
XX		
KW	Human; Plexin; semaphorin domain; h	
KW	neurodegenerative disease; autoimmu	
KW	inflammatory bowel disease; diabete	
KW	immunogen; antibody.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200114420-A2.	
XX		
PD	01-MAR-2001.	
XX		
PF	25-AUG-2000; 2000WO-US23365.	
XX		
PR	25-AUG-1999; 99US-0150576.	
XX		
PA	(UYTO-) UNIV TORINO.	
PA	(REGC ) UNIV CALIFORNIA.	

RESULT	14	
AAU00020		
ID	AAU00020	standard; Protein; 729 AA.
XX		
AC	AAU00020;	
XX		
DT	09-MAY-2001	(first entry)
XX		
DE	Human Plexin protein.	
XX		
KW	Human; Plexin; semaphorin domain; hyperplasia; neoplasia; cancer;	
KW	neurodegenerative disease; autoimmune disease; lupus; multiple sclerosis;	
KW	inflammatory bowel disease; diabetes type I; rheumatoid arthritis;	
KW	immunogen; antibody.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200114420-A2.	
XX		
PD	01-MAR-2001.	
XX		
PF	25-AUG-2000; 2000WO-US23365.	
XX		
PR	25-AUG-1999; 99US-0150576.	
XX		
PA	(UYTO-) UNIV TORINO.	
PA	(REGC ) UNIV CALIFORNIA.	



XX Artigiani S, Comoglio PM, Goodman CS, Tesier-Lavigne M;  
PI Tamagnone L;  
XX WPI; 2001-226610/23.  
DR N-PSDB; AAS00025.  
XX New plexin polynucleotides and polypeptides, useful in diagnosis,  
PT therapy and in producing compounds for treating diseases involving  
PT aberrant cell growth (e.g. cancer) or immune regulation (e.g.  
PT autoimmune diseases)  
XX Disclosure; Page 78-79; 79pp; English.  
PS The sequence represents a Human Plexin. Plexins are large transmembrane  
XX proteins whose extracellular domain shares homology with scatter factor  
CC receptors and contain an approximately 500 amino acid Semaphorin domain.  
CC The plexin polynucleotides and polypeptides, and the plexin-specific binding  
CC agents are useful in diagnosis, therapy and in the biopharmaceutical  
CC industry. In particular, the plexin polynucleotides and polypeptides are  
CC useful for generating compounds (e.g. plexin-specific binding agents or  
CC antibodies) for treating or diagnosing a disease or disorder involving  
CC aberrant cell growth (e.g. hyperplasia, neoplasia, cancer or  
CC neurodegenerative disease), or diseases or disorders involving aberrant  
CC immune regulation (e.g. autoimmune diseases such as lupus, inflammatory  
CC bowel disease or Diabetes Type I), or immunosuppressive diseases such as  
XX multiple sclerosis or rheumatoid arthritis.  
SQ Sequence 729 AA;

Query Match 60.3%; Score 41; DB 22; Length 729;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPLE 12  
Db 584 sgvmcspdpse 595  
:|::||| | | |

RESULT 15  
AAB21203  
ID AAB21203 standard; protein; 919 AA.

XX AAB21203;

XX 12-JAN-2001 (first entry)

XX Human HER-2/neu fusion protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer; fusion protein.

XX Homo sapiens.  
OS Synthetic.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.  
PA (SMIK ) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX WPI; 2000-505976/45.

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins

PT useful for vaccinating against breast, ovarian, colon, lung and  
PI prostate cancers -  
XX Claim 2; Fig 12; 128pp; English.  
XX The present sequence is a fusion protein comprising the extracellular  
CC domain and the phosphorylation domain of the human HER-2/neu protein.  
CC HER-2/neu is a member of the tyrosine kinase family of receptor-like  
CC glycoproteins and shows homology to the epidermal growth factor receptor  
CC (EGFR). It probably plays a part in cell growth and/or differentiation.  
CC The HER-2/neu gene is an oncogene. HER-2/neu fusion proteins may be used  
CC to treat or prevent cancer by eliciting or enhancing an immune response  
CC to the HER-2/neu protein. They may be used to treat malignancies such as  
CC breast, ovarian, colon, lung and prostate cancers, and may be used as an  
CC antigen to vaccinate against these neoplasias.  
XX Sequence 919 AA;

Query Match 60.3%; Score 41; DB 21; Length 919;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
Db 693 gffcpdpap 701  
| | | | |

RESULT 16

AAB21208  
ID AAB21208 standard; Protein; 1200 AA.

XX AAB21208;

XX 12-JAN-2001 (first entry)

XX Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.

XX Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.  
PA (SMIK ) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX WPI; 2000-505976/45.

XX N-PSDB; AAA89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -

XX Disclosure; Fig 15; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member  
CC of the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or

CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against  
CC these neoplasias.

XX SQ Sequence 1200 AA;

Query Match 60.3%; Score 41; DB 21; Length 1200;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| | | | |  
Db 1029 gffcpdpap 1037

RESULT 17  
AAW01111  
ID AAW01111 standard; Protein; 1255 AA.  
XX  
AC AAW01111;  
XX  
DT 01-JAN-1997 (first entry)  
XX  
DE HER-2/neu protein.  
XX  
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;  
KW breast cancer; ovary cancer; colon cancer; lung cancer;  
KW prostate cancer; immunisation; tumour; vaccine; vector.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 676..1255  
FT /label= Intracellular\_domain  
FT /note= "claimed domain, useful for immunisation"

XX PN WO9630514-A1.

XX PD 03-OCT-1996.

XX PF 28-MAR-1996; 96WO-US01689.

XX PR 31-MAR-1995; 95US-0414417.

XX PA (UNIW ) UNIV WASHINGTON.

XX PI Cheever MA, Disis ML;

XX DR WPI; 1996-455361/45.

XX DR N-PSDB; AAT40739.

XX PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or  
PT treatment of malignancies with which the HER-2/neu oncogene is  
PT associated  
XX  
PS Claim 2; Page 56-61; 71pp; English.

XX CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AAT40739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transformed host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 60.3%; Score 41; DB 17; Length 1255;

Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| | | | |  
Db 1029 gffcpdpap 1037

RESULT 18  
AAW92406  
ID AAW92406 standard; Protein; 1255 AA.  
XX  
AC AAW92406;

XX DT 21-APR-1999 (first entry)  
XX  
DE Human HER-2/neu oncogene protein.

XX KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
KW malignancy; treatment; tumour.  
XX  
OS Homo sapiens.

XX FH Key Location/Qualifiers  
FT Region 676..1255  
FT /note= "region which elicits immune response"

XX PN US5869445-A.

XX PD 09-FEB-1999.

XX PF 01-APR-1996; 96US-0625101.

XX PR 01-APR-1996; 96US-0625101.

XX PR 17-MAR-1993; 93US-0033644.

XX PR 12-AUG-1993; 93US-0106112.

XX PR 31-MAR-1995; 95US-0414417.

XX PA (UNIW ) UNIV WASHINGTON.

XX PI Cheever MA, Disis ML;

XX DR WPI; 1999-152835/13.

XX DR N-PSDB; AAX01912.

XX PT Use of HER-2/neu polypeptides - for eliciting an immune response to  
PT an HER-2/neu associated malignancy, particularly for treating or  
PT preventing tumours

XX PS Claim 3; Column 31-38; 26pp; English.

XX CC This sequence represents the human HER-2/neu oncogene protein. A fragment  
CC of this protein is used in a method for eliciting or enhancing an immune  
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
CC B cells to produce an immune response to the HER-2/neu protein. The  
CC method can be used for immunisation against a malignancy in which the  
CC HER-2/neu oncogene is associated and in the treatment of an existing  
CC tumour, or to prevent tumour occurrence or reoccurrence.

XX SQ Sequence 1255 AA;

Query Match 60.3%; Score 41; DB 20; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| | | | |  
Db 1029 gffcpdpap 1037

RESULT 19  
AAB21198

ID AAB21198 standard; protein; 1255 AA.  
XX AAB21198;  
AC  
XX  
XX  
DT 12-JAN-2001 (first entry)  
XX  
XX  
DE Human HER-2/neu protein.  
XX  
XX  
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200044899-A1.  
XX  
XX  
PD 03-AUG-2000.  
XX  
XX  
PF 28-JAN-2000; 2000WO-US02164.  
XX  
XX  
PR 29-JAN-1999; 99US-0117976.  
XX  
XX  
PA (CORI-) CORIXA CORP.  
PA (SMIK ) SMITHKLINE BEECHAM.  
XX  
XX  
PI Cheever MA, Gheysen D;  
XX  
XX  
DR WPI; 2000-505976/45.  
DR N-PSDB; AAA89736.  
XX  
XX  
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX  
XX  
PS Claim 52; Fig 7; 128pp; English.  
XX  
XX  
CC The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against  
CC these neoplasias.  
XX  
XX  
SQ Sequence 1255 AA;  
  
Query Match 60.3%; Score 41; DB 21; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 GLLCPDPRP 10  
DB 1029 gffcpdpap 1037  
  
RESULT 20  
AAY84780  
ID AAY84780 standard; Protein; 1255 AA.  
XX  
XX  
AC AAY84780;  
XX  
XX  
DT 08-AUG-2000 (first entry)  
XX  
XX  
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.  
XX  
XX  
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;  
KW tumor cell proliferation; tissue degeneration; arthropathy;  
KW bone resorption; inflammatory disease; degenerative disorder;

KW wound healing.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200020579-A1.  
XX  
XX  
PD 13-APR-2000.  
XX  
XX  
PF 01-OCT-1999; 99WO-CA00912.  
XX  
XX  
PR 02-OCT-1998; 98US-0165192.  
XX  
XX  
PA (UYMC-) UNIV MCMASTER.  
XX  
XX  
PI Muller WJ, Siegel PM;  
XX  
XX  
DR WPI; 2000-303768/26.  
DR N-PSDB; AAA14812.  
XX  
XX  
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE  
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -  
XX  
XX  
PS Claim 3; Fig 2; 60pp; English.  
XX  
XX  
CC The present sequence represents a SPLICE erbB-2 receptor protein. The  
CC protein has an in-frame deletion of 16 amino acids, 2 of which are  
CC conserved cysteine residues, compared to the unspliced protein. The  
CC erbB-2 polynucleotide is used to construct probes for detecting  
CC disorders of cell transformation such as cancer. Antibodies to the  
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents  
CC (e.g. antisense oligonucleotides) which inhibit the expression of  
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and  
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful  
CC for treating conditions involving damaged cells including conditions  
CC in which degeneration of tissue occurs, such as arthropathy, bone  
CC resorption, inflammatory diseases, degenerative disorders of the  
CC central nervous system and wound healing.  
XX  
XX  
SQ Sequence 1255 AA;  
  
Query Match 60.3%; Score 41; DB 21; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 GLLCPDPRP 10  
DB 1029 gffcpdpap 1037  
  
RESULT 21  
AAY92620  
ID AAY92620 standard; Protein; 1255 AA.  
XX  
XX  
AC AAY92620;  
XX  
XX  
DT 10-AUG-2000 (first entry)  
XX  
XX  
DE Human heregulin 2 (Her2).  
XX  
XX  
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;  
KW self-protein; cancer; breast cancer; prostate cancer;  
KW cell-associated peptide antigen; foreign epitope.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key  
FH Domain  
FT Location/Qualifiers  
FT 1..173  
FT /label= N-terminal  
FT /note= "mature polypeptide"  
FT 5..25  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"

FT	Region	59..73	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	103..117	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	149..163	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Domain	174..323	
FT		/label= Cysteine_rich_domain	
FT	Region	210..224	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	250..264	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Domain	324..483	
FT		/label= Ligand_binding_domain	
FT	Region	325..339	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	369..383	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	465..479	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Domain	484..623	
FT		/label= Cysteine_rich_domain	
FT	Region	579..593	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Domain	624..654	
FT		/label= Transmembrane_domain	
FT	Region	632..652	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	653..667	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Domain	655..1010	
FT		/label= Tyrosine_kinase_domain	
FT	Region	661..675	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	695..709	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Region	710..730	
FT		/label= insertion_region	
FT		/note= "suitable for foreign epitope insertion"	
FT	Domain	1011..1235	
FT		/label= C-terminal_domain	
XX			
PN	WO200020027-A2.		
XX			
PD	13-APR-2000.		
XX			
PF	05-OCT-1999;	99WO-DK00525.	
XX			
PR	05-OCT-1998;	98DK-0001261.	
PR	20-OCT-1998;	98US-0105011.	
XX			
PA	(MEBI-) M & E BIOTECH AS.		
XX			
PI	Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Da		
PI	Gautam A, Birk P, Karlsson G;		
XX			
DR	WPI; 2000-349917/30.		
DR	N-PSDB; AAA09455.		
XX			
PT	Inducing immune responses to weakly immunoqenic, tumor associat		

peptide analogues for the treatment of breast and prostate cancer

Claim 62; Page 193-198; 220pp; English.

This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of Her2 can be used in the claimed method as an autovaccine to induce a CTL response. Subdominant CTL epitopes, antibody binding regions and cysteine residues involved in disulfide bonds are preserved in the immunogenized forms. Regions suitable for the insertion of foreign T helper epitopes were identified (see features table). The method is used for inducing immune responses against weakly immunogenic cell-associated peptide antigens (PA) such as those associated with cancers (self-proteins), e.g. human prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b). The method comprises effecting simultaneous presentation by antigen producing cells (APCs) of the animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1 B-cell group derived from the cell-associated PA; and (2) at least 1 first T helper cell group which is foreign to the animal. Analogues of human PSM, human Her2 and human/murine FGF8b comprising a substantial part of all known and predicted CTL and B-cell epitopes of the respective PA and including at least one foreign T helper epitope are also claimed. The method is used to treat prostate, prostate/breast or breast cancer when the PA is human PSM, FGF8b and Her2, respectively.

Sequence 1255 AA;

Query Match 60.3%; Score 41; DB 21; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
I I I I I I  
Db 1029 gffcpdpap 1037

RESULT 22  
AAB60167  
ID AAB60167 standard; Protein; 1255 AA.  
XX AAB60167;  
AC  
XX 03-APR-2001 (first entry)  
XX  
DE HER2 transgene plasmid construct encoded protein.  
XX  
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;  
KW antibody.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200100244-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 23-JUN-2000; 2000WO-US17229.  
PF  
XX  
PR 25-JUN-1999; 99US-0141316.  
PR 16-MAR-2000; 2000US-0189844.  
XX  
XX (GETH ) GENENTECH INC.  
PA  
XX  
XX Erickson S, Schwall R;  
PI  
XX  
DR WPI; 2001-061962/07.  
DR N-PSDB; AAF24297.  
XX  
PT Treating tumors, particularly breast cancers, which overexpress an ErbB  
PT receptor and does not respond to an anti-ErbB antibody, comprises  
PT conjugating the antibody to a maytansinoid -  
XX



PS Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by

CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In

CC particular, the antibody is directed against ErbB2 (also known as HER2

CC and p185neu). The method is particularly useful in the treatment of

CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,

CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX Sequence 1255 AA;

QY 2 GLLCPDPRP 10

Db 1029 gffcpdpap 1037

RESULT 23

AAR39568

ID AAR39568 standard; Protein; 1433 AA.

XX AC

XX AAR39568;

XX DT

XX 07-FEB-1994 (first entry)

XX DE

XX Sequence of c-erbB-2 tumour antigen.

XX KW

XX Tumour antigen; c-erbB-2; glycoprotein.

XX OS

XX Homo sapiens.

XX PN

XX WO9316185-A.

XX PD

XX 19-AUG-1993.

XX PF

XX 05-FEB-1993; 93WO-US01055.

XX PR

XX 06-FEB-1992; 92US-0831967.

XX XX

XX (CETU ) CETUS ONCOLOGY CORP.

XX PA

XX (CREA-) CREATIVE BIOMOLECULES INC.

XX PI

XX Houston LL, Huston JS, Oppermann H, Ring DB;

XX DR

XX WPI; 1993-272889/34.

XX DR

XX N-PSDB; AAQ46083.

XX XX

XX New single chain Fv polypeptide binding to C-erbB-2 tumour

XX PT

XX antigen - for imaging or treating breast or ovarian cancer etc.

XX XX

XX Disclosure; pages 48-54; 87pp; English.

XX PS

XX c-erbB-2 refers to a protein antigen expressed on the surface of

XX CC

XX tumour cells. such as breast and ovarian tumour cells, which is an

XX CC

XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

XX CC

XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents

XX CC

XX the location of a stop codon in AAQ46083.

XX XX

XX Sequence 1433 AA;

QY 2 GLLCPDPRP 10

Db 1029 gffcpdpap 1037

Query Match 60.3%; Score 41; DB 14; Length 1433;

Best Local Similarity 66.7%; Pred. No. 2.5e+02;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

PS Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by

CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In

CC particular, the antibody is directed against ErbB2 (also known as HER2

CC and p185neu). The method is particularly useful in the treatment of

CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,

CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

XX Sequence 1255 AA;

QY 2 GLLCPDPRP 10

Db 1029 gffcpdpap 1037

RESULT 24

AAU00019

ID AAU00019 standard; Protein; 2135 AA.

XX AC

XX AAU00019;

XX DT

XX 09-MAY-2001 (first entry)

XX DE

XX Human Plexin protein.

XX KW

XX Human; Plexin; semaphorin domain; hyperplasia; neoplasia; cancer;

XX KW

XX neurodegenerative disease; autoimmune disease; lupus; multiple sclerosis;

XX KW

XX inflammatory bowel disease; diabetes type I; rheumatoid arthritis;

XX XX

XX immunogen; antibody.

XX OS

XX Homo sapiens.

XX FH

XX Key

XX Location/Qualifiers

XX FT

XX Misc-difference 1297 /note= "Encoded by AAC"

XX FT

XX Misc-difference 1625 /note= "Encoded by ACC"

XX PN

XX WO200114420-A2.

XX XX

XX 01-MAR-2001.

XX XX

XX 25-AUG-2000; 2000WO-US23365.

XX PF

XX 25-AUG-1999; 99US-0150576.

XX PR

XX (UYTO-) UNIV TORINO.

XX PA

XX (REGC ) UNIV CALIFORNIA.

XX XX

XX Artigiani S, Comoglio PM, Goodman CS, Tesier-Lavigne M;

XX PI

XX Tamagnone L;

XX XX

XX WPI; 2001-226610/23.

XX DR

XX N-PSDB; AAS00024.

XX XX

XX New plexin polynucleotides and polypeptides, useful in diagnosis,

XX PT

XX therapy and in producing compounds for treating diseases involving

XX PT

XX aberrant cell growth (e.g. cancer) or immune regulation (e.g.

XX PT

XX autoimmune diseases)

XX XX

XX Disclosure; Page 72-77; 79pp; English.

XX PS

XX The sequence represents a Human Plexin. Plexins are large transmembrane

XX CC

XX proteins whose extracellular domain shares homology with Scatter factor

XX CC

XX receptors and contain an approximately 500 amino acid semaphorin domain.

XX CC

XX The plexin polynucleotides and polypeptides, and the biopharmaceutical

XX CC

XX agents are useful in diagnosis, therapy and in the biopharmaceutical

XX CC

XX industry. In particular, the plexin polynucleotides and polypeptides are

XX CC

XX useful for generating compounds (e.g. plexin-specific binding agents or

XX CC

XX antibodies) for treating or diagnosing a disease or disorder involving

XX CC

XX aberrant cell growth (e.g. hyperplasia, neoplasia, cancer or

XX CC

XX neurodegenerative disease), or diseases or disorders involving aberrant

XX CC

XX immune regulation (e.g. autoimmune diseases such as lupus, inflammatory

XX CC

XX bowel disease or Diabetes type I), or immunosuppressive diseases such as

XX CC

XX multiple sclerosis or rheumatoid arthritis.

XX XX

XX Sequence 2135 AA;

QY 1 AGLLCPDPRPLE 12

Db 584 sgvmcspdpse 595

Query Match 60.3%; Score 41; DB 22; Length 2135;

Best Local Similarity 50.0%; Pred. NO. 3.6e+02;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

RESULT 25  
AAR15105  
ID AAR15105 standard; Protein; 122 AA.  
XX  
XX  
AC AAR15105;  
DT 11-FEB-1992 (first entry)  
XX  
DE hCG/bLH chimera, D9.  
XX  
KW Glycoprotein hormone; immuno-castration;  
KW immuno-contragestive; vaccine; human chorionic gonadotropin;  
KW luteinising hormone; LH; CG; bovine.  
XX  
OS Homo sapiens.  
OS Bos taurus.  
XX  
PN WO9116922-A.  
XX  
PD 14-NOV-1991.  
XX  
PF 07-MAY-1991; 91WO-US03162.  
XX  
PR 08-MAY-1990; 90US-0520703.  
XX  
PA (UYNE-) UNIV MED NEW JERSEY.  
XX  
PI Campbell RK, Moyle WR;  
XX  
DR WPI; 1991-353528/48.  
XX  
XX  
PT New glyco-protein hormone analogues - for inducing fertility as  
PT immuno-castration agents, for suppressing reproductive system  
PT development and as immuno-contragestive vaccines.  
XX  
PS Table IV; Page 63; 94pp; English.  
XX  
CC The sequence is an analogue of mature hCG beta subunit having  
CC several residues replaced by the corresponding residues in the  
CC bovine LH protein and residues 23-145 deleted. The chimeric hormone  
CC may be useful for identifying residues which are important for  
CC binding to the human receptor and may also have applications as an  
CC immunogen, agonist and/or antagonist.  
CC See AAR15043, AAR15061-R15125 and AAR15161-R15198.  
XX  
SQ Sequence 122 AA;

Query Match 58.8%; Score 40; DB 12; Length 122;  
Best Local Similarity 77.8%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11  
| | | | |  
Db 108 ltcddprpl 116

RESULT 26  
AAW05504  
ID AAW05504 standard; Protein; 328 AA.  
XX  
XX  
AC AAW05504;  
XX  
DT 15-JAN-1997 (first entry)  
XX  
DE HCMV Toledo strain UL135 protein (clone tol.03).  
XX  
KW CMV; HCMV; vaccine; diagnosis; UL135.  
XX  
OS Human cytomegalovirus Toledo strain.  
XX  
PN WO9630387-A1.

XX  
PD 03-OCT-1996.  
XX  
PF 26-MAR-1996; 96WO-US04100.  
XX  
PR 31-MAR-1995; 95US-0414926.  
XX  
PA (AVIR-) AVIRON.  
XX  
PI Cha T, Spaete R;  
XX  
DR WPI; 1996-455265/45.  
DR N-PSDB; AAT41418.  
XX  
PT New isolated human cytomegalovirus nucleic acid - from Towne and  
PT Toledo strains, used to develop prods. for the diagnosis, prevention  
PT and treatment of human CMV infections  
XX  
PS Claim 5; Page 69-70; 150pp; English.  
XX  
CC Novel protein UL135 (AAW05504) is the product of an open reading  
CC frame found in a novel nucleic acid (AAT41418) isolated from the  
CC Toledo strain of human cytomegalovirus (HCMV). UL135 and other  
CC novel (see also AAW05502-20) and known (see also AAW05500-01) proteins  
CC of the Toledo strain, as well as new proteins (see also AAW05521-24)  
CC from HCMV Towne, can be produced in transformed host cells and used  
CC in the prodn. of subunit vaccines against HCMV. They may be  
CC surface glycoproteins that are immunogenic or responsible for  
CC tissue tropism, or may influence the immune response of an infected  
CC individual.  
XX  
SQ Sequence 328 AA;

Query Match 58.8%; Score 40; DB 17; Length 328;  
Best Local Similarity 77.8%; Pred. No. 94;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| | | | |  
Db 273 glscpcprp 281

RESULT 27  
AAG31001  
ID AAG31001 standard; Protein; 485 AA.  
XX  
XX  
AC AAG31001;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37160.

XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.

PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 58.8%; Score 40; DB 21; Length 485;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9  
|:|||||  
Db 375 licpdpr 381

RESULT 28  
AAG31000  
ID AAG31000 standard; Protein; 486 AA.  
XX  
AC AAG31000;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37159.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.



PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0148368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 58.8%; Score 40; DB 21; Length 486;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 LLCPPDR 9  
Db 376 licpdpr 382

RESULT 29  
AAY38566  
ID AAY38566 standard; Protein: 522 AA.  
XX  
AC AAY38566;  
XX  
DT 08-OCT-1999 (first entry)  
XX  
DE Neisseria meningitidis antigen encoded by ORF12.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9924578-A2.  
XX  
PD 20-MAY-1999.  
XX  
PF 09-OCT-1998; 98WO-IB01665.  
XX  
PR 01-SEP-1998; 98GB-0019016.  
PR 06-NOV-1997; 97GB-0023516.  
PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR 27-NOV-1997; 97GB-0025158.  
PR 10-DEC-1997; 97GB-0026147.  
PR 14-JAN-1998; 98GB-0000759.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
XX  
DR WPI; 1999-327407/27.  
DR N-PSDB; AAZ12030.  
XX  
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
XX diagnosis, treatment and prevention of infection  
PS Claim 4; Page 129; 524pp; English.  
XX  
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,

CC their fragments, their nucleic acids and antibodies are used for  
CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
CC infections, such as meningitis, septicaemia and gonorrhea. Both  
CC organisms are closely related. Fragments of the nucleic acids  
CC are useful as hybridisation probes and antisense reagents.

XX Sequence 522 AA;

Query Match 58.8%; Score 40; DB 20; Length 522;  
Best Local Similarity 70.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GLLCDPRPL 11  
|| |||||  
Db 48 glsvdprrpv 57

RESULT 30  
AAY38568  
ID AAY38568 standard; Protein; 522 AA.  
XX  
AC AAY38568;  
XX  
DT 08-OCT-1999 (first entry)  
XX  
DE Neisseria gonorrhoeae antigen encoded by ORF12.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
XX  
OS Neisseria gonorrhoeae.  
XX  
PN WO9924578-A2.  
XX  
PD 20-MAY-1999.  
XX  
PF 09-OCT-1998; 98WO-IB01665.  
XX  
PR 01-SEP-1998; 98GB-0019016.  
PR 06-NOV-1997; 97GB-0023516.  
PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR 27-NOV-1997; 97GB-0025158.  
PR 10-DEC-1997; 97GB-0026147.  
PR 14-JAN-1998; 98GB-0000759.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;  
XX  
DR WPI; 1999-327407/27.  
DR N-PSDB; AAZ12032.  
XX  
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
PT diagnosis, treatment and prevention of infection  
XX  
PS Claim 4; Page 132; 524pp; English.  
XX  
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,  
CC their fragments, their nucleic acids and antibodies are used for  
CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
CC infections, such as meningitis, septicaemia and gonorrhea. Both  
CC organisms are closely related. Fragments of the nucleic acids  
CC are useful as hybridisation probes and antisense reagents.

XX Sequence 522 AA;

Query Match 58.8%; Score 40; DB 20; Length 522;  
Best Local Similarity 70.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 GLLCDPRPL 11  
|| |||||  
Db 48 glsvdprrpv 57

RESULT 31  
AAY38567  
ID AAY38567 standard; Protein; 523 AA.  
XX  
AC AAY38567;  
XX  
DT 08-OCT-1999 (first entry)  
XX  
DE Neisseria meningitidis strain A antigen encoded by ORF12.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9924578-A2.  
XX  
PD 20-MAY-1999.  
XX  
PF 09-OCT-1998; 98WO-IB01665.  
XX  
PR 01-SEP-1998; 98GB-0019016.  
PR 06-NOV-1997; 97GB-0023516.  
PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR 27-NOV-1997; 97GB-0025158.  
PR 10-DEC-1997; 97GB-0026147.  
PR 14-JAN-1998; 98GB-0000759.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;  
XX  
DR WPI; 1999-327407/27.  
DR N-PSDB; AAZ12031.  
XX  
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
PT diagnosis, treatment and prevention of infection  
XX  
PS Claim 4; Page 130; 524pp; English.  
XX  
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,  
CC their fragments, their nucleic acids and antibodies are used for  
CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
CC infections, such as meningitis, septicaemia and gonorrhea. Both  
CC organisms are closely related. Fragments of the nucleic acids  
CC are useful as hybridisation probes and antisense reagents.

XX Sequence 523 AA;

Query Match 58.8%; Score 40; DB 20; Length 523;  
Best Local Similarity 70.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GLLCDPRPL 11  
|| |||||  
Db 49 glsvdprrpv 58

RESULT 32  
AAY74891  
ID AAY74891 standard; Protein; 526 AA.  
XX

AC AAY74891;  
DT 21-MAR-2000 (first entry)  
DE Neisseria meningitidis ORF 302 protein sequence SEQ ID NO:1256.  
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
DE antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy.  
KW Neisseria meningitidis.  
OS WO9957280-A2.  
PN 11-NOV-1999.  
PD 30-APR-1999; 99WO-US093346.  
PF 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX WPI; 2000-062150/05.  
DR N-PSDB; AAZ53653.  
XX Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics  
PT Claim 2; Page 696; 1453pp; English.  
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
XX represent novel Neisseria meningitis and N. gonorrhoea polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX SQ Sequence 526 AA;  
Query Match 58.8%; Score 40; DB 21; Length 526;  
Best Local Similarity 70.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GLLCPDPRPL 11  
Db 59 glsvdpdprpv 68  
RESULT 33  
AAY74890  
ID AAY74890 standard; Protein; 533 AA.  
XX AAY74890;  
AC

XX 21-MAR-2000 (first entry)  
DT Neisseria gonorrhoea ORF 302 protein sequence SEQ ID NO:1254.  
DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy.  
KW Neisseria gonorrhoeae.  
OS WO9957280-A2.  
PN 11-NOV-1999.  
PD 30-APR-1999; 99WO-US093346.  
PF 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX WPI; 2000-062150/05.  
DR N-PSDB; AAZ53652.  
XX Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics  
PT Claim 2; Page 695; 1453pp; English.  
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
XX represent novel Neisseria meningitis and N. gonorrhoea polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX SQ Sequence 533 AA;  
Query Match 58.8%; Score 40; DB 21; Length 533;  
Best Local Similarity 70.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GLLCPDPRPL 11  
Db 59 glsvdpdprpv 68  
RESULT 34  
AAY74892  
ID AAY74892 standard; Protein; 533 AA.  
XX AAY74892;  
AC

DT 21-MAR-2000 (first entry)  
XX Neisseria meningitidis ORF 302 protein sequence SEQ ID NO:1258.  
DE  
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy.  
KW  
XX Neisseria meningitidis.  
OS  
XX WO9957280-A2.  
PN  
XX 11-NOV-1999.  
PD  
XX 30-APR-1999; 99WO-US09346.  
PF  
XX 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;  
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
DR WPI; 2000-062150/05.  
DR N-PSDB; AAZ53654.  
XX  
PT Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics -  
XX  
PS Claim 2; Page 697-698; 1453pp; English.  
XX  
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAY75941  
CC represent novel Neisseria meningitis and N. gonorrhoea polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX  
SQ Sequence 533 AA;

Query Match 58.8%; Score 40; DB 21; Length 533;  
Best Local Similarity 70.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPPRPL 11  
|| |||||  
Db 59 glsvdprrpv 68

RESULT 35  
AAG30999  
ID AAG30999 standard; Protein; 538 AA.  
XX  
AC AAG30999;  
XX  
DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 37158.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
PN  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.



PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 25-JUN-1999; 99US-0140695.  
PR 26-JUN-1999; 99US-0140823.  
PR 27-JUN-1999; 99US-0140991.  
PR 28-JUN-1999; 99US-0141287.  
PR 29-JUN-1999; 99US-0141842.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 03-JUL-1999; 99US-0142390.  
PR 04-JUL-1999; 99US-0142803.  
PR 05-JUL-1999; 99US-0142920.  
PR 06-JUL-1999; 99US-0142977.  
PR 07-JUL-1999; 99US-0143542.  
PR 08-JUL-1999; 99US-0143624.  
PR 09-JUL-1999; 99US-0144005.  
PR 10-JUL-1999; 99US-0144085.  
PR 11-JUL-1999; 99US-0144086.  
PR 12-JUL-1999; 99US-0144325.  
PR 13-JUL-1999; 99US-0144331.  
PR 14-JUL-1999; 99US-0144332.  
PR 15-JUL-1999; 99US-0144333.  
PR 16-JUL-1999; 99US-0144334.  
PR 17-JUL-1999; 99US-0144335.  
PR 18-JUL-1999; 99US-0144352.  
PR 19-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 22-JUL-1999; 99US-0144814.  
PR 23-JUL-1999; 99US-0145086.  
PR 24-JUL-1999; 99US-0145088.  
PR 25-JUL-1999; 99US-0145088.  
PR 26-JUL-1999; 99US-0145087.  
PR 27-JUL-1999; 99US-0145089.  
PR 28-JUL-1999; 99US-0145192.  
PR 29-JUL-1999; 99US-0145145.  
PR 30-JUL-1999; 99US-0145218.  
PR 01-AUG-1999; 99US-0145224.  
PR 02-AUG-1999; 99US-0145276.  
PR 03-AUG-1999; 99US-0145913.  
PR 04-AUG-1999; 99US-0145918.  
PR 05-AUG-1999; 99US-0145919.  
PR 06-AUG-1999; 99US-0145951.  
PR 07-AUG-1999; 99US-0146386.  
PR 08-AUG-1999; 99US-0146388.  
PR 09-AUG-1999; 99US-0146389.  
PR 10-AUG-1999; 99US-0147038.  
PR 11-AUG-1999; 99US-0147204.  
PR 12-AUG-1999; 99US-0147302.  
PR 13-AUG-1999; 99US-0147192.  
PR 14-AUG-1999; 99US-0147260.  
PR 15-AUG-1999; 99US-0147303.  
PR 16-AUG-1999; 99US-0147416.  
PR 17-AUG-1999; 99US-0147935.  
PR 18-AUG-1999; 99US-0148171.  
PR 19-AUG-1999; 99US-0148319.  
PR 20-AUG-1999; 99US-0148341.  
PR 21-AUG-1999; 99US-0148565.  
PR 22-AUG-1999; 99US-0148684.  
PR 23-AUG-1999; 99US-0149368.  
PR 24-AUG-1999; 99US-0149175.  
PR 25-AUG-1999; 99US-0149426.  
PR 26-AUG-1999; 99US-0149722.  
PR 27-AUG-1999; 99US-0149723.  
PR 28-AUG-1999; 99US-0149929.  
PR 29-AUG-1999; 99US-0149902.  
PR 30-AUG-1999; 99US-0149930.  
PR 01-SEP-1999; 99US-0150566.  
PR 02-SEP-1999; 99US-0150884.  
PR 03-SEP-1999; 99US-0151065.  
PR 04-SEP-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.  
PR 28-AUG-1999; 99US-0151303.  
PR 29-AUG-1999; 99US-0151438.  
PR 30-AUG-1999; 99US-0151930.  
PR 31-AUG-1999; 99US-0152363.  
PR 01-SEP-1999; 99US-0153070.  
PR 02-SEP-1999; 99US-0153758.  
PR 03-SEP-1999; 99US-0154018.  
PR 04-SEP-1999; 99US-0154039.  
PR 05-SEP-1999; 99US-0154779.  
PR 06-SEP-1999; 99US-0155139.  
PR 07-SEP-1999; 99US-0155486.  
PR 08-SEP-1999; 99US-0155659.  
PR 09-SEP-1999; 99US-0156458.  
PR 10-SEP-1999; 99US-0156596.  
PR 11-SEP-1999; 99US-0157117.  
PR 12-SEP-1999; 99US-0157753.  
PR 13-SEP-1999; 99US-0157865.  
PR 14-SEP-1999; 99US-0158029.  
PR 15-SEP-1999; 99US-0158232.  
PR 16-SEP-1999; 99US-0158369.  
PR 17-SEP-1999; 99US-0159293.  
PR 18-SEP-1999; 99US-0159294.  
PR 19-SEP-1999; 99US-0159295.  
PR 20-SEP-1999; 99US-0159329.  
PR 21-SEP-1999; 99US-0159330.  
PR 22-SEP-1999; 99US-0159331.  
PR 23-SEP-1999; 99US-0159637.  
PR 24-SEP-1999; 99US-0159638.  
PR 25-SEP-1999; 99US-0159584.  
PR 26-SEP-1999; 99US-0160741.  
PR 27-SEP-1999; 99US-0160767.  
PR 28-SEP-1999; 99US-0160768.  
PR 29-SEP-1999; 99US-0160770.  
PR 30-SEP-1999; 99US-0160814.  
PR 01-OCT-1999; 99US-0160815.  
PR 02-OCT-1999; 99US-0160980.  
PR 03-OCT-1999; 99US-0160981.  
PR 04-OCT-1999; 99US-0160989.  
PR 05-OCT-1999; 99US-0161404.  
PR 06-OCT-1999; 99US-0161405.  
PR 07-OCT-1999; 99US-0161406.  
PR 08-OCT-1999; 99US-0161359.  
PR 09-OCT-1999; 99US-0161360.  
PR 10-OCT-1999; 99US-0161361.  
PR 11-OCT-1999; 99US-0161920.  
PR 12-OCT-1999; 99US-0161992.  
PR 13-OCT-1999; 99US-0161993.  
PR 14-OCT-1999; 99US-0162142.

Query Match 58.8%; Score 40; DB 21; Length 538;  
Best Local Similarity 85.7%; Pred No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDP 9  
Db 428 licpdpr 434

RESULT 36  
AAG22586  
ID AAG22586 standard; Protein; 161 AA.  
XX  
AC AAG22586;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 25572.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX

OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999;  
PR 01-JUL-1999;  
PR 01-JUL-1999;  
PR 02-JUL-1999;  
PR 06-JUL-1999;  
PR 08-JUL-1999;  
PR 09-JUL-1999;  
PR 12-JUL-1999;  
PR 13-JUL-1999;  
PR 14-JUL-1999;  
PR 15-JUL-1999;  
PR 16-JUL-1999;  
PR 16-JUL-1999;  
PR 19-JUL-1999;  
PR 19-JUL-1999;  
PR 19-JUL-1999;  
PR 19-JUL-1999;  
PR 19-JUL-1999;  
PR 20-JUL-1999;  
PR 20-JUL-1999;  
PR 20-JUL-1999;  
PR 21-JUL-1999;  
PR 21-JUL-1999;  
PR 21-JUL-1999;  
PR 22-JUL-1999;  
PR 22-JUL-1999;  
PR 22-JUL-1999;  
PR 22-JUL-1999;  
PR 23-JUL-1999;  
PR 23-JUL-1999;  
PR 26-JUL-1999;  
PR 27-JUL-1999;  
PR 27-JUL-1999;  
PR 27-JUL-1999;  
PR 28-JUL-1999;  
PR 02-AUG-1999;  
PR 02-AUG-1999;  
PR 02-AUG-1999;  
PR 03-AUG-1999;  
PR 04-AUG-1999;  
PR 04-AUG-1999;  
PR 05-AUG-1999;  
PR 05-AUG-1999;  
PR 06-AUG-1999;  
PR 06-AUG-1999;  
PR 09-AUG-1999;  
PR 09-AUG-1999;  
PR 10-AUG-1999;  
PR 11-AUG-1999;  
PR 12-AUG-1999;  
PR 13-AUG-1999;  
PR 13-AUG-1999;  
PR 16-AUG-1999;  
PR 17-AUG-1999;  
PR 18-AUG-1999;  
PR 20-AUG-1999;  
PR 20-AUG-1999;  
PR 20-AUG-1999;  
PR 23-AUG-1999;  
PR 23-AUG-1999;  
PR 25-AUG-1999;  
PR 26-AUG-1999;  
PR 27-AUG-1999;  
PR 27-AUG-1999;  
PR 27-AUG-1999;  
PR 30-AUG-1999;  
PR 31-AUG-1999;  
PR 01-SEP-1999;  
PR 07-SEP-1999;  
PR 10-SEP-1999;  
PR 13-SEP-1999;

99US-0141287.  
99US-0141842.  
99US-0142154.  
99US-0142055.  
99US-0142390.  
99US-0142803.  
99US-0142920.  
99US-0142977.  
99US-0143542.  
99US-0143624.  
99US-0144005.  
99US-0144085.  
99US-0144086.  
99US-0144325.  
99US-0144331.  
99US-0144332.  
99US-0144333.  
99US-0144334.  
99US-0144335.  
99US-0144352.  
99US-0144632.  
99US-0144884.  
99US-0144814.  
99US-0145086.  
99US-0145088.  
99US-0145085.  
99US-0145087.  
99US-0145089.  
99US-0145192.  
99US-0145145.  
99US-0145218.  
99US-0145224.  
99US-0145276.  
99US-0145913.  
99US-0145918.  
99US-0145919.  
99US-0145951.  
99US-0146386.  
99US-0146388.  
99US-0146389.  
99US-0147038.  
99US-0147204.  
99US-0147302.  
99US-0147192.  
99US-0147260.  
99US-0147303.  
99US-0147416.  
99US-0147493.  
99US-0147935.  
99US-0148171.  
99US-0148319.  
99US-0148341.  
99US-0148565.  
99US-0148684.  
99US-0149175.  
99US-0149368.  
99US-0149175.  
99US-0149426.  
99US-0149722.  
99US-0149723.  
99US-0149929.  
99US-0149902.  
99US-0149930.  
99US-0150566.  
99US-0150884.  
99US-0151065.  
99US-0151066.  
99US-0151080.  
99US-0151303.  
99US-0151438.  
99US-0151930.  
99US-0152363.  
99US-0153070.  
99US-0153758.

PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 57.4%; Score 39; DB 21; Length 161;  
Best Local Similarity 85.7%; Pred. NO: 70;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9  
Db 78 lvcppdr 84

RESULT 37  
AAG33891  
ID AAG33891 standard; Protein; 172 AA.  
XX AC  
AC AAG33891;  
XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 41151.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX OS  
XX EP1033405-A2.  
XX PN  
XX 06-SEP-2000.  
XX PD  
XX 25-FEB-2000; 2000EP-0301439.  
XX PF

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 57.4%; Score 39; DB 21; Length 172;  
Best Local Similarity 85.7%; Pred. No. 74;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9  
I:|||||  
Db 110 lvcpdpr 116

RESULT 38  
AAY18084  
ID AAY18084 standard; Protein; 198 AA.

XX AC AAY18084;  
XX 06-AUG-1999 (first entry)  
DT DT  
XX Histamine binding protein Av-HBP.  
DE Histamine binding protein; serotonin binding compound; inflammation;  
XX Histamine binding protein; allergy; type I hypersensitivity reaction;  
KW gastric acid secretion; allergic rhinitis; atopic dermatitis; food allergy;  
KW asthma; urticaria; allergic rhinitis; atopic dermatitis; food allergy;  
KW drug allergy; abnormal blood pressure; psychological disorder; vaccine;  
KW respiratory disease; coronary heart disease; cellular growth regulator;  
KW tissue repair; blood-sucking ectoparasite; therapy.

OS Amblyomma variegatum.  
XX WO9927104-A1.  
PN 03-JUN-1999.  
XX 26-NOV-1998; 98WO-GB03530.  
PF 26-JUN-1998; 98GB-0013917.  
XX 26-NOV-1997; 97GB-0025046.  
XX



PA (OXFO-) OXFORD VACS LTD.  
XX  
PI Nuttall PA, Paesen GC;  
XX  
DR WPI; 1999-357841/30.  
XX N-PSDB; AAX76969.  
XX  
XX Histamine and serotonin binding compounds useful for the treatment  
PT of allergies  
XX  
XX Claim 12; Fig 6; 84pp; English.  
XX  
XX This sequence is an example of a histamine or serotonin binding  
CC compound (A), of the invention. The compounds are useful for regulating  
CC the action of histamine and serotonin (in e.g. inflammation and gastric  
CC acid secretion), the detection, quantification and removal of histamine  
CC or serotonin (in animals, plants, cell cultures, food materials, or  
CC humans) and in the treatment of various diseases and allergies  
CC (e.g. type I hypersensitivity reactions, urticaria, asthma, allergic  
CC rhinitis (hay fever), atopic dermatitis, insect bites and food and drug  
CC allergies, abnormal blood pressure, migraine, psychological disorders,  
CC respiratory disease, and coronary heart disease). Histamine may also be  
CC used to regulate cellular growth and tissue repair. The molecules may  
CC also be used as components of vaccines directed against blood-sucking  
CC ectoparasites.  
XX  
SQ Sequence 198 AA;

Query Match 57.4%; Score 39; DB 20; Length 198;  
Best Local Similarity 85.7%; Pred. NO. 84;  
Matches 6; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 4 LCPDPRP 10  
Db 177 lcpqprp 183

RESULT 39  
AAB27731  
ID AAB27731 standard; Protein; 264 AA.  
XX  
XX AAB27731;  
XX  
XX 30-JAN-2001 (first entry)  
XX  
XX Sequence homologous to protein fragment encoded by gene 2.  
DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX  
XX Homo sapiens.  
OS  
XX WO200055201-A1.  
PN  
XX 21-SEP-2000.  
PD  
XX 09-MAR-2000; 2000WO-US06059.  
PF  
XX 12-MAR-1999; 99US-0124096.  
XX 03-DEC-1999; 99US-0168622.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Rosen CA, Ruben SM, Komatsoulis G;  
XX  
XX WPI; 2000-628182/60.  
XX  
XX Novel human secreted proteins useful for diagnosis, prevention and  
PT treatment of disorders including neurological, cell proliferative,

PT cardiovascular, autoimmune/inflammatory disorders and microbial  
XX infections  
XX  
XX Disclosure; Page 9; 427pp; English.  
XX  
XX The invention relates to the isolation of genes AAC59157-C59205 encoding  
CC the human secreted proteins AAB27682-B27730. This sequence represents a  
CC peptide fragment homologous to the protein encoded by the gene given  
CC in the descriptor line. The sequence is a search result from a BLASTX  
CC homology search. The genes and proteins are useful for preventing,  
CC ameliorating or treating medical conditions, e.g. by protein or gene  
CC therapy. The genes are isolated from a range of human tissues disclosed  
CC in the specification. The nucleic acids, proteins, antibodies and  
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such  
CC as myocardial ischaemia; (d) wound healing; (e) neurological diseases  
CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such  
CC as viral, bacterial, fungal and parasitic infections.  
XX  
SQ Sequence 264 AA;

Query Match 57.4%; Score 39; DB 21; Length 264;  
Best Local Similarity 85.7%; Pred. NO. 1.le+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CPDPRPL 11  
Db 1 cpdppl 7

RESULT 40  
AAW77301  
ID AAW77301 standard; Protein; 285 AA.  
XX  
XX AAW77301;  
XX  
XX 07-JAN-1999 (first entry)  
XX  
XX Amino acid sequence of murine HELA2.  
DE  
XX Serine protease; regulation; cell activity; viability; HELA2; ATC2;  
KW BCOM3; testisin; fertility; suppressor; testicular germ cell cancer;  
KW seminoma; testis-specific expression; antitumour; sperm development;  
KW infertility; mouse.  
XX  
XX Mus sp.  
OS  
XX WO9836054-A1.  
PN  
XX 20-AUG-1998.  
PD  
XX 13-FEB-1998; 98WO-AU00085.  
XX  
XX 18-NOV-1997; 97AU-0000422.  
XX 13-FEB-1997; 97AU-0005101.  
PR  
XX (AMRA-) AMRAD OPERATIONS PTY LTD.  
PA  
XX Antalis TM, Hooper JD;  
XX  
XX WPI; 1998-480768/41.  
XX N-PSDB; AAV59132.  
DR  
XX New serine protease(s) and kinase involved in regulating cell  
PT activity and viability - particularly the testis-specific protease  
PT HELA2 used for modulation of fertility and as tumour suppressor  
XX

PS Example 14; Fig 18A; 167pp; English.

XX The present sequence represents the amino acid sequence of murine HELA2.

CC Human HELA2 was isolated from Hela cells. HELA2 has high homology to

CC serine proteases. The protein is involved in or associated with

CC regulation of cell activity and/or viability. Administration of

CC recombinant HELA2 (also called testisin) is used to increase fertility.

CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of

CC testicular germ cell cancers (seminoma) and is also expressed in some

CC non-testicular cancers (of colon, pancreas, prostate and ovary), so is

CC a marker/potential therapeutic target for cancer. The promoter from the

CC HELA2 gene is useful for testis-specific expression of other genes,

CC e.g. for gene therapy or modulation of fertility. Drugs that block

CC activity of HELA2 should have antitumour activity (other than in

CC testis) while in testis recombinant HELA2 should stop growth of tumours

CC and normalise sperm development (eliminating the need for orchidectomy).

CC Identification of mutant forms of HELA2 can be used to diagnose

CC infertility.

XX

SQ Sequence 285 AA;

Query Match 57.4%; Score 39; DB 19; Length 285;

Best Local Similarity 80.0%; Pred. No. 1.2e+02;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCDPDRPL 11.

Db 260 gllrpdvpvl 269

RESULT 41

AAY59390

ID AAY59390 standard; Protein; 379 AA.

XX

AC AAY59390;

XX

DT 14-MAR-2000 (first entry)

XX

DE Murine soluble interleukin-11 receptor.

XX

KW Interleukin-11; IL-11; IL-11R; human; bone density disorder; gp130;

KW IL-11 tertiary complex; glycoprotein 130; postmenopausal bone loss;

KW bone resorption inhibitor; bone formation; therapy.

XX

OS Mus sp.

XX

PN WO9959608-A2.

XX

PD 25-NOV-1999.

XX

PF 19-MAY-1999; 99WO-CA00516.

XX

PR 19-MAY-1998; 98CA-2237915.

XX

PA (HAMI-) HAMILTON CIVIC HOSPITAL RES DEV CORP.

XX

PI Shaughnessy S, Austin RC;

XX

DR WPI; 2000-062377/05.

DR N-PSDB; AAZ40400.

XX

PT Inhibiting formation of a tertiary complex for the treatment of

PT osteoporosis -

XX

PS Example 3; Page 46-50; 61pp; English.

XX

CC This sequence is the soluble mouse interleukin-11 receptor (IL-11R).

CC The invention relates to a method of treating or alleviating the symptoms

CC of a pathological condition in which bone density is decreased comprises

CC inhibiting the formation of a tertiary complex of IL-11, IL-11 receptor

CC and glycoprotein 130 (gp130) in a mammalian patient suffering from such a

CC condition. The method is used to treat or alleviate the symptoms of a

CC pathological condition in which bone density is decreased, especially

CC postmenopausal bone loss. The IL-11 binding peptide is useful in the

CC purification of IL-11 or in depleting IL-11 from a solution. TRAP

CC (tartrate-resistant acid phosphatase) and bone marrow formation assays

CC can be used for the identification of IL-11 antagonists. The method not

CC only inhibits bone resorption and hence bone loss, but also increases the

CC process of bone formation to increase bone density.

XX

SQ Sequence 379 AA;

Query Match 57.4%; Score 39; DB 21; Length 379;

Best Local Similarity 77.8%; Pred. No. 1.5e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPLE 12

Db 352 lqpdprpld 360

RESULT 42

AAB42988

ID AAB42988 standard; Protein; 392 AA.

XX

AC AAB42988;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF2752 polypeptide sequence SEQ ID NO:5504.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US08621.

XX

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach M;

XX

DR WPI; 2000-602362/57.

DR N-PSDB; AAC77197.

XX

PT Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease -

XX

PS Claim 11; Page 4686-4687; 5507pp; English.

XX

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 392 AA;

Query Match 57.4%; Score 39; DB 21; Length 392;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLLCPDPR 9  
DB 57 agrlcpppr 65

RESULT 43  
AAR92813  
ID AAR92813 standard; Protein; 432 AA.

XX AC AAR92813;

XX DT 21-MAY-1996 (first entry)

XX DE Murine interleukin-11 receptor alpha chain.

XX DE Haemopoietin; interleukin-11; IL-11; receptor; agonist; antagonist;  
KW therapy; diagnosis.

XX OS Mus sp.

XX FH Key Location/Qualifiers  
FT Peptide 1..23  
FT Protein /label= Sig\_peptide  
FT Domain 24..432  
FT Domain /label= Mat\_protein  
FT Domain 24..367

FT /label= Extracellular\_domain  
FT /note= "the extracellular domain includes  
FT haemopoietin and Ig-like domains"

FT Domain 368..393

FT Domain /label= Transmembrane\_domain

FT Domain 394..432

FT /label= Cytoplasmic\_tail

XX WO9607737-A1.

XX DT 14-MAR-1996.

XX PF 05-SEP-1995; 95WO-AU00578.

XX PR 05-SEP-1994; 94AU-0007902.

XX PR 05-SEP-1994; 94AU-0007901.

XX PA (AMRA-) AMRAD OPERATIONS PTY LTD.

XX PI Hilton DJ;

XX WPI; 1996-171612/17.  
DR N-PSDB; AAT17868.  
XX Nucleic acid encoding haemopoietin receptor containing conserved  
PT amino acid motif esp. IL-11 receptor alpha chain - used for  
PT developing IL-11 (ant)agonists  
XX Claim 6; Page 42-44; 87pp; English.  
XX The murine interleukin-11 (IL-11) receptor alpha chain Nrl  
CC (AAR92813) was identified by expression of DNA (AAT17868) isolated  
CC from adult mouse liver cDNA libraries. Nrl is a low affinity  
CC receptor for IL-11 and interacts with gp130 to generate a high  
CC affinity IL-11 receptor. IL-11 can be used to develop  
CC agonists or antagonists of therapeutic appln. or in the treatment  
CC or diagnosis of conditions involving a deficiency of IL-11,  
CC excess IL-11 or aberrant effects of normal endogenous IL-11  
XX levels.  
XX Sequence 432 AA;

Query Match 57.4%; Score 39; DB 17; Length 432;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPLE 12  
DB 352 lqpdprpld 360

RESULT 44  
AAB36653  
ID AAB36653 standard; Protein; 432 AA.

XX AC AAB36653;

XX DT 13-MAR-2001 (first entry)

XX DE Mouse IL-11 receptor subunit alpha protein SEQ ID NO:10.

XX DE DNAX cytokine receptor subunit; DCRS2; receptor protein;  
KW modulating cell proliferation; diagnosis; detection; drug screening;  
KW immunological disorder.

XX OS Mus sp.

XX PN WO200073451-A1.

XX PD 07-DEC-2000.

XX PF 30-MAY-2000; 2000WO-US14867.

XX PR 01-JUN-1999; 99US-0322913.

XX PA (SCHE ) SCHERING CORP.

XX PI Dowling LM, Timans JC, Gorman DM, Kastelein RA, Bazan FJ;

XX WPI; 2001-061536/07.

XX Novel composition comprising DNAX cytokine receptor subunit polypeptide  
PT useful for regulating immune system function and for treating  
PT immunological disorders

XX PS Disclosure; Page 13-15; 93pp; English.

XX The present invention describes a composition (I) comprising a  
CC recombinant DNAX cytokine receptor subunit-2 (DCRS2) polypeptide.  
CC The DCRS2 polypeptide is useful for binding ligands and for preparing  
CC antibodies. The DCRS2 polypeptide is also useful for modulating cell  
CC proliferation, for diagnostic and therapeutic applications, for

CC detecting presence of their ligands and in drug screening assays. It  
CC is also useful for treating conditions such as immunological disorders.  
CC The present sequence represents a cytokine receptor subunit protein  
CC which is given in an alignment of various cytokine receptor subunits in  
CC the exemplification of the present invention.

XX  
SQ Sequence 432 AA;

Query Match 57.4%; Score 39; DB 22; Length 432;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPLE 12  
| | | | |  
Db 352 lqpdprpld 360

RESULT 45  
AAR99091  
ID AAR99091 standard; Protein; 441 AA.  
XX  
AC AAR99091;  
XX  
DT 09-OCT-1996 (first entry)  
XX  
DE Murine Etl-2 gene product.  
XX  
KW Interleukin-11 receptor; IL-11 receptor; antibody; antagonist; bone;  
KW osteoporosis; Paget disease; myeloma; Etl-2.  
XX  
OS Mus sp.  
XX  
PN WO9619574-A1.  
XX  
PD 27-JUN-1996.  
XX  
PF 27-NOV-1995; 95WO-US15400.  
XX  
PR 22-DEC-1994; 94US-0362304.  
XX  
PA (GEMY ) GENETICS INST INC.  
XX  
PI Tobin JF;  
XX  
DR WPI; 1996-309588/31.  
DR N-PSDB; AAT32613.  
XX  
PT New nucleic acid encoding human interleukin 11 receptor - and  
PT related protein, antibodies, receptor antagonists, etc, useful for  
PT treating and preventing loss of bone mass  
XX  
PS Example 1; Page 37-40; 54pp; English.  
XX

CC The amino acid sequence of the murine Etl-2 gene product is given  
CC in AAR99091. Probes based on the Etl-2 gene (AAT32613) were used  
CC to screen a human cDNA library, yielding a cDNA clone (AAT33278) that  
CC coded for human interleukin-11 receptor (AAR99090), a protein  
CC involved in bone maturation and repair.  
XX  
SQ Sequence 441 AA;

Query Match 57.4%; Score 39; DB 17; Length 441;  
Best Local Similarity 77.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPLE 12  
| | | | |  
Db 361 lqpdprpld 369

Search completed: June 28, 2001, 11:54:14  
Job time: 75 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:09 ; Search time 37.02 Seconds  
(without alignments)  
6.530 Million cell updates/sec

Title: US-09-439-313-558  
Perfect score: 68  
Sequence: 1 AGLICPDRPLE 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCRTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	63.2	362	2	US-09-055-097-7
2	41	60.3	580	1	US-08-414-417B-69
3	41	60.3	580	2	US-08-486-348A-69
4	41	60.3	580	2	US-08-468-545B-69
5	41	60.3	580	3	US-08-466-680B-69
6	41	60.3	1255	1	US-08-467-083-68
7	41	60.3	1255	1	US-08-414-417B-68
8	41	60.3	1255	2	US-08-484-438-8
9	41	60.3	1255	2	US-08-486-348A-68
10	41	60.3	1255	2	US-08-625-101-2
11	41	60.3	1255	2	US-08-468-545B-68
12	41	60.3	1255	2	US-08-356-786-2
13	41	60.3	1255	3	US-08-466-680B-68
14	40	58.8	328	1	US-08-414-926A-9
15	40	58.8	328	2	US-08-926-922-9
16	40	58.8	328	3	US-09-253-682-9
17	39.5	58.1	595	2	US-08-677-049-11
18	38	55.9	190	1	US-08-441-629-4
19	38	55.9	190	3	US-08-776-207-4
20	38	55.9	190	5	PCT-US95-09172-4
21	38	55.9	999	2	US-08-473-553A-5
22	38	55.9	2595	4	US-09-036-987A-2
23	37	54.4	279	1	US-08-300-903A-9
24	37	54.4	617	4	US-09-188-930-303
25	37	54.4	1290	1	US-08-470-350B-2
26	37	54.4	3072	4	US-09-413-814-93
27	37	54.4	3079	4	US-09-413-814-80

28	36	52.9	96	1	US-08-414-926A-11	Sequence 11, Appl
29	36	52.9	96	2	US-08-926-922-11	Sequence 11, Appl
30	36	52.9	96	3	US-09-253-682-11	Sequence 11, Appl
31	36	52.9	233	1	US-08-221-750A-7	Sequence 7, Appli
32	36	52.9	1052	2	US-08-852-806-2	Sequence 2, Appli
33	36	52.9	1052	4	US-09-163-669-2	Sequence 2, Appli
34	35	51.5	288	3	US-08-934-494-4	Sequence 4, Appli
35	35	51.5	288	3	US-09-143-068-4	Sequence 5, Appli
36	35	51.5	732	1	US-08-317-522A-5	Sequence 5, Appli
37	35	51.5	778	1	US-08-439-818A-5	Sequence 5, Appli
38	35	51.5	778	2	US-08-751-965-5	Sequence 5, Appli
39	35	51.5	778	2	US-08-738-975-5	Sequence 5, Appli
40	35	51.5	778	2	US-08-728-626-5	Sequence 5, Appli
41	35	51.5	778	4	US-08-808-599A-5	Sequence 5, Appli
42	35	51.5	4472	2	US-08-804-227C-2	Sequence 2, Appli
43	34	50.0	60	2	US-08-465-794-14	Sequence 14, Appl
44	34	50.0	60	3	US-09-049-813-14	Sequence 14, Appl
45	34	50.0	80	4	US-08-663-191A-1	Sequence 1, Appli
46	34	50.0	80	4	US-09-051-624A-1	Sequence 1, Appli
47	34	50.0	95	1	US-07-891-942G-16	Sequence 16, Appl
48	34	50.0	157	2	US-08-599-602-2	Sequence 2, Appli
49	34	50.0	157	3	US-09-197-816-2	Sequence 2, Appli
50	34	50.0	178	2	US-08-465-794-18	Sequence 18, Appl
51	34	50.0	178	3	US-09-049-813-18	Sequence 18, Appl
52	34	50.0	178	4	US-08-663-191A-4	Sequence 4, Appli
53	34	50.0	189	4	US-09-369-494-15	Sequence 15, Appl
54	34	50.0	447	4	US-09-305-001-2	Sequence 2, Appli
55	34	50.0	521	2	US-08-504-048-9	Sequence 9, Appli
56	34	50.0	521	5	PCT-US93-08386-10	Sequence 10, Appl
57	34	50.0	548	3	US-08-656-034-10	Sequence 10, Appl
58	34	50.0	628	3	US-08-656-034-2	Sequence 2, Appli
59	34	50.0	1079	3	US-09-058-489-22	Sequence 22, Appl
60	34	50.0	1240	3	US-09-058-489-23	Sequence 23, Appl
61	34	50.0	1347	3	US-09-058-489-24	Sequence 24, Appl
62	33	48.5	51	4	US-08-905-223-384	Sequence 384, App
63	33	48.5	146	3	US-08-476-120-8	Sequence 8, Appli
64	33	48.5	225	1	US-08-300-903A-14	Sequence 14, Appl
65	33	48.5	234	1	US-08-300-903A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-055-097-7  
; Sequence 7, Application US/09055097  
; Patent No. 5955282  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; APPLICANT: Patterson, Chandra  
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/055,097  
; FILING DATE: Filed Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1469908
; US-09-055-097-7

Query Match      63.2%; Score 43; DB 2; Length 362;
Best Local Similarity 87.5%; Pred. NO. 13;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11
Db 92 LCPDPSPL 99

RESULT 2
US-08-414-417B-69
; Sequence 69, Application US/08414417B
; Patent No. 5801005
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,417B
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C2
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-414-417B-69

Query Match      60.3%; Score 41; DB 1; Length 580;

```

```

Best Local Similarity 66.7%; Pred. NO. 42;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10
Db 354 GFFCPDPAP 362

RESULT 3
US-08-486-348A-69
; Sequence 69, Application US/08486348A
; Patent No. 5846538
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-486-348A-69

Query Match      60.3%; Score 41; DB 2; Length 580;
Best Local Similarity 66.7%; Pred. NO. 42;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10
Db 354 GFFCPDPAP 362

RESULT 4
US-08-468-545B-69
; Sequence 69, Application US/08468545B
; Patent No. 5876712
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP

```

STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,545B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 580 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-468-545B-69

Query Match 60.3%; Score 41; DB 2; Length 580;  
Best Local Similarity 66.7%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| |||||  
Db 354 GFFCPDPAP 362

RESULT 5

US-08-466-680B-69  
Sequence 69, Application US/084666680B  
Patent No. 6075122  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,680B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 580 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-466-680B-69

Query Match 60.3%; Score 41; DB 3; Length 580;  
Best Local Similarity 66.7%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| |||||  
Db 354 GFFCPDPAP 362

RESULT 6

US-08-467-083-68  
Sequence 68, Application US/08467083  
Patent No. 5726023  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN  
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
TITLE OF INVENTION: HER-2/NEU ONCOGENE IS ASSOCIATED  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,083  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/414,417  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836 SEEDANBERRY  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-467-083-68

Query Match 60.3%; Score 41; DB 1; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| |||||  
Db 1029 GFFCPDPAP 1037



```

RESULT 7
US-08-414-417B-68
; Sequence 68, Application US/08414417B
; Patent No. 5801005
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/414,417B
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-414-417B-68

```

```

Query Match 60.3%; Score 41; DB 1; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 GLLCPDPRP 10
DB 1029 GFFCPDPAP 1037

```

```

RESULT 8
US-08-484-438-8
; Sequence 8, Application US/084844438
; Patent No. 5811098
; Patent No. 5811098 5780031
; GENERAL INFORMATION:
; APPLICANT: Plowman, Gregory D.
; APPLICANT: Culouscou, Jean-Michel
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Siegall, Clay B.
; APPLICANT: Hellstr m, Ingegerd
; APPLICANT: Hellstr m, Karl E.
; TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,438
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA: 08/323,442
; APPLICATION NUMBER: 08/323,442
; FILING DATE: 14-OCT-1994
; APPLICATION NUMBER: US 08/150,704
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/981,165
; FILING DATE: 24-NOV-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: protein
; MOLECULE TYPE: protein
; US-08-484-438-8

```

```

Query Match 60.3%; Score 41; DB 2; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 GLLCPDPRP 10
DB 1029 GFFCPDPAP 1037

```

```

RESULT 9
US-08-486-348A-68
; Sequence 68, Application US/08486348A
; Patent No. 5846538
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

```

NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-486-348A-68

Query Match 60.3%; Score 41; DB 2; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10  
I I I I I I  
Db 1029 GFFCPDPAP 1037

RESULT 10  
US-08-625-101-2  
Sequence 2, Application US/08625101  
Patent No. 5869445  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE  
REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION  
TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu  
TITLE OF INVENTION: ONCOGENE IS ASSOCIATED  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/625,101  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C7  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-625-101-2

Query Match 60.3%; Score 41; DB 2; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10  
I I I I I I

Db 1029 GFFCPDPAP 1037

RESULT 11  
US-08-468-545B-68  
Sequence 68, Application US/08468545B  
Patent No. 5876712  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,545B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-468-545B-68

Query Match 60.3%; Score 41; DB 2; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10  
I I I I I I  
Db 1029 GFFCPDPAP 1037

RESULT 12  
US-08-356-786-2  
Sequence 2, Application US/08356786  
Patent No. 5877305  
GENERAL INFORMATION:  
APPLICANT: Huston, James S.  
APPLICANT: Oppermann, Hermann  
APPLICANT: Houston, L. L.  
APPLICANT: Ring, David B.  
TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault  
STREET: Exchange Place, 53 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,786
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/831,967
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-356-786-2

Query Match 60.3%; Score 41; DB 2; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10
Db 1029 GFFCPDPAP 1037

RESULT 13
US-08-466-680B-68
; Sequence 68, Application US/08466680B
; Patent No. 6075122
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,680B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:

```

```

;
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-466-680B-68

Query Match 60.3%; Score 41; DB 3; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10
Db 1029 GFFCPDPAP 1037

RESULT 14
US-08-414-926A-9
; Sequence 9, Application US/08414926A
; Patent No. 5721354
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,926A
; FILING DATE: March 31, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; IMMEDIATE SOURCE:
; CLONE: tol.03
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..328
; OTHER INFORMATION: /label= ULL35
; US-08-414-926A-9

Query Match 58.8%; Score 40; DB 1; Length 328;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10
Db 273 GLSCPCPRP 281

RESULT 15
US-08-926-922-9

```

```

; Sequence 9, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.03
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..328
; OTHER INFORMATION: /label= UL135
; US-08-926-922-9

Query Match 58.8%; Score 40; DB 2; Length 328;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10
   || || ||
Db 273 GLSCPCRP 281

RESULT 16
US-09-253-682-9
; Sequence 9, Application US/09253682
; Patent No. 6040170
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cserr Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/253,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/926,922
; APPLICATION NUMBER: 08/926,922
; FILING DATE: September 10, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.03
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..328
; OTHER INFORMATION: /label= UL135
; US-09-253-682-9

Query Match 58.8%; Score 40; DB 3; Length 328;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10
   || || ||
Db 273 GLSCPCRP 281

RESULT 17
US-08-677-049-11
; Sequence 11, Application US/08677049
; Patent No. 5858707
; GENERAL INFORMATION:
; APPLICANT: Guimaraes, M. Jorge
; APPLICANT: Bazan, J. Fernando
; APPLICANT: McClanahan, Terrill K.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
; NUCLEIC ACIDS; ANTIBODIES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,049
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,788
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.

```



```
;
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Region
; LOCATION: 233-262
; OTHER INFORMATION: /note= "Encompasses TM 4 of Figure
; OTHER INFORMATION: 4"
;
; NAME/KEY: Region
; LOCATION: 459-521
; OTHER INFORMATION: /note= "Encompasses TM 9 and TM 10
; OTHER INFORMATION: Of Figure 4"
;
US-08-677-049-11

Query Match 58.1%; Score 39.5; DB 2; Length 595;
Best Local Similarity 61.5%; Pred. No. 73;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 3;
QY 2 GLLCPD---PRPL 11
Db 273 GMLCPSATAPRPL 285

RESULT 18
US-08-441-629-4
; Sequence 4, Application US/08441629
; Patent No. 5766923
; GENERAL INFORMATION:
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Kinoshita, No. 5766923iyuki
; TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,629
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,217
; FILING DATE: 22-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: HU95-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
```

```
;
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-441-629-4

Query Match 55.9%; Score 38; DB 1; Length 190;
Best Local Similarity 72.7%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 GLLCPDPRPLE 12
Db 28 GLNCNDPRLE 38

RESULT 19
US-08-776-207-4
; Sequence 4, Application US/08776207A
; Patent No. 6080718
; GENERAL INFORMATION:
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Kinoshita, No. 6080718iyuki
; TITLE OF INVENTION: Receptor-Ligand Assay
; FILE REFERENCE: HU95-01A2
; CURRENT APPLICATION NUMBER: US/08/776,207A
; CURRENT FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: PCT/US95/09172
; EARLIER FILING DATE: 1995-07-19
; EARLIER APPLICATION NUMBER: 08/441,629
; EARLIER FILING DATE: 1995-05-15
; EARLIER APPLICATION NUMBER: 08/279,217
; EARLIER FILING DATE: 1994-07-22
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Xenopus laevis
; US-08-776-207-4

Query Match 55.9%; Score 38; DB 3; Length 190;
Best Local Similarity 72.7%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 GLLCPDPRPLE 12
Db 28 GLNCNDPRLE 38

RESULT 20
PCT-US95-09172-4
; Sequence 4, Application PC/TUS9509172
; GENERAL INFORMATION:
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Kinoshita, Noriyuki
; TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09172
; FILING DATE:
```

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/279,217  
FILING DATE: 22-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/441,629  
FILING DATE: 15-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: HU95-01A PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-09172-4

Query Match 55.9%; Score 38; DB 5; Length 190;  
Best Local Similarity 72.7%; Pred. No. 41;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDRPLE 12  
||| |||||  
Db 28 GLNCNDPRLE 38

RESULT 21

US-08-473-553A-5  
Sequence 5, Application US/08473553A  
Patent No. 5859338  
GENERAL INFORMATION:  
APPLICANT: Meyerowitz, Elliot M.  
APPLICANT: Clark, Steven E.  
APPLICANT: Williams, Robert W.  
TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
TITLE OF INVENTION: Transformed Plants, and Proteins  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,553A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 999 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown

TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-473-553A-5

Query Match 55.9%; Score 38; DB 2; Length 999;  
Best Local Similarity 70.0%; Pred. No. 2.1e+02;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDRPL 11  
|||||  
Db 943 GLLCTSPPL 952

RESULT 22

US-09-036-987A-2  
Sequence 2, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
TITLE OF INVENTION: Production  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow AgroSciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036,987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-036-987A-2

Query Match 55.9%; Score 38; DB 4; Length 2595;  
Best Local Similarity 72.7%; Pred. No. 5.4e+02;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLICPDRPL 11  
|||||  
Db 2390 AGLLAARPL 2400

RESULT 23  
US-08-300-903A-9

```

; Sequence 9, Application US/08300903A
; Patent No. 5591630
; GENERAL INFORMATION:
; APPLICANT: Anderson, Dirk M
; APPLICANT: Giri, Judith G
; TITLE OF INVENTION: Interleukin-15 Receptors
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Operating System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICANT: Snyder, Solomon H
; TITLE OF INVENTION: Ebnerein: A Secreted von Ebner's Gland
; NUMBER OF SEQUENCES: 6
; FILING DATE: 06-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/236,919
; FILING DATE: 06-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne
; REGISTRATION NUMBER: 34,695
; REFERENCE/DOCKET NUMBER: 2822-A
; TELEPHONE: 206-587-0430
; TELEFAX: 206-233-0644
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-300-903A-9

```

```

Query Match      54.4%; Score 37; DB 1; Length 279;
Best Local Similarity 50.0%; Pred. No: 85;
Matches 6; Conservative 2; Mismatches 4; Indels 4; Gaps 0;

```

```

QY 1 AGLLCPDPRPLE 12
   ||: || | : |
Db 74 AGITCPPPMSE 85

```

```

RESULT 24
US-09-188-930-303
; Sequence 303, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 303
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Mouse

```

US-09-188-930-303

```

Query Match      54.4%; Score 37; DB 4; Length 617;
Best Local Similarity 77.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 3 LILCPDPRPL 11
   ||| ||||
Db 520 LILCAAPRPL 528

```

```

RESULT 25
US-08-470-350B-2
; Sequence 2, Application US/08470350B
; Patent No. 5684126
; GENERAL INFORMATION:
; APPLICANT: Li, Xiao
; APPLICANT: Snyder, Solomon H
; TITLE OF INVENTION: Ebnerein: A Secreted von Ebner's Gland
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,350B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wolfe, Susan A
; REGISTRATION NUMBER: 33,568
; REFERENCE/DOCKET NUMBER: 01107.48790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1290 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-350B-2

```

```

Query Match      54.4%; Score 37; DB 1; Length 1290;
Best Local Similarity 58.3%; Pred. No. 3.9e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

```

```

QY 1 AGLLC--PDPRP 10
   ||: || |
Db 811 AGVICSLPDPTP 822

```

```

RESULT 26
US-09-413-814-93
; Sequence 93, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra

```

```

; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 93
; LENGTH: 3072
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-93

Query Match          54.4%; Score 37; DB 4; Length 3072;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AGLLCPDPR 9
||: || ||
Db 44 AGVSCPSR 52

RESULT 27
US-09-413-814-80
; Sequence 80, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 3079
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-80

Query Match          54.4%; Score 37; DB 4; Length 3079;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AGLLCPDPR 9
||: || ||
Db 47 AGVSCPSR 55

RESULT 28

```

```

US-08-414-926A-11
; Sequence 11, Application US/08414926A
; Patent No. 5721354
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,926A
; FILING DATE: March 31, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.05
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..96
; OTHER INFORMATION: /label= UL137
US-08-414-926A-11

Query Match          52.9%; Score 36; DB 1; Length 96;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 LLCPPDRP 10
:|||||
Db 26 VLCPDLRP 33

RESULT 29
US-08-926-922-11
; Sequence 11, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cseri Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```





QY 6 PDRPL 11  
|||||  
Db 2 PDRPL 7

RESULT 32

US-08-852-806-2  
; Sequence 2, Application US/08852806  
; Patent No. 5874245

GENERAL INFORMATION:

APPLICANT: Shoji Fukusumi  
APPLICANT: Shuji Hinuma  
TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA

COUNTRY: U.S.A.

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 1.5  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/852,806  
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/017,915

FILING DATE: 16 MAY 1996

ATTORNEY/AGENT INFORMATION:

NAME: Han, William T

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: TAK50002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1052 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

Query Match 52.9%; Score 36; DB 2; Length 1052;  
Best Local Similarity 66.7%; Pred. No. 4.5e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPDPRP 10  
||: ||||  
Db 325 GLILPPRP 333

RESULT 33

US-09-163-669-2

; Sequence 2, Application US/09163669  
; Patent No. 6111076

GENERAL INFORMATION:

APPLICANT: FUKUSUMI, SHOJI

APPLICANT: HINUMA, SHUJI

APPLICANT: FUJII, RYO

TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED

; TITLE OF INVENTION: RECEPTOR (HIBCD07)  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Ratner & Prestia

STREET: P.O. Box 980

CITY: Valley Forge

STATE: PA

COUNTRY: USA

ZIP: 19482

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/163,669

FILING DATE: 30-SEP-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/852,806

FILING DATE: 07-MAY-1997

APPLICATION NUMBER: 60/017,915

FILING DATE: 16-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Prestia, Paul F

REGISTRATION NUMBER: 23,031

REFERENCE/DOCKET NUMBER: TAK-50002-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-407-0700

TELEFAX: 610-407-0700

TELEX: 846169

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1052 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-163-669-2

Query Match 52.9%; Score 36; DB 4; Length 1052;  
Best Local Similarity 66.7%; Pred. No. 4.5e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPDPRP 10  
||: ||||  
Db 325 GLILPPRP 333

RESULT 34

US-08-934-494-4

; Sequence 4, Application US/08934494  
; Patent No. 6030831

GENERAL INFORMATION:

APPLICANT: Gurney, Austin

TITLE OF INVENTION: Tie Ligands

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/934,494

FILING DATE:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1078  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-3216  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-934-494-4

Query Match 51.5%; Score 35; DB 3; Length 288;  
Best Local Similarity 75.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CPDPRPLE 12  
Db 29 CPGPRELE 36

RESULT 35  
US-09-143-068-4  
Sequence 4, Application US/09143068  
Patent No. 6074873  
GENERAL INFORMATION:  
APPLICANT: Fong, Sherman  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth  
APPLICANT: Williams, Mickey  
TITLE OF INVENTION: Tie Ligand Homologues  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/143,068  
FILING DATE: 28-Aug-1998

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1078P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-3216  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 288 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-143-068-4

Query Match 51.5%; Score 35; DB 3; Length 288;  
Best Local Similarity 75.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CPDPRPLE 12  
Db 29 CPGPRELE 36

RESULT 36  
US-08-317-522A-5  
Sequence 5, Application US/08317522A  
Patent No. 5599918  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
TITLE OF INVENTION: Trophinin and Trophinin-Assisting Proteins  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,522A  
FILING DATE: 04-OCT-1994

CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 9991  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-522A-5

Query Match 51.5%; Score 35; DB 1; Length 732;  
Best Local Similarity 70.0%; Pred. No. 4.5e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12  
Db 581 CPPAEPRLPE 590

RESULT 37  
US-08-439-818A-5  
Sequence 5, Application US/08439818A  
Patent No. 5654145  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
TITLE OF INVENTION: Trophinin and Trophinin-Assisting Proteins  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/439,818A  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,522  
; FILING DATE: 04-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 1563  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 778 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-439-818A-5

Query Match 51.5%; Score 35; DB 1; Length 778;  
Best Local Similarity 70.0%; Pred. No. 4.8e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12  
|| :|||||  
Db 581 CPPAEPRLPLE 590

RESULT 38  
US-08-751-965-5  
; Sequence 5, Application US/08751965  
; Patent No. 5858360  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/751,965  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 12-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 778 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-751-965-5

Query Match 51.5%; Score 35; DB 2; Length 778;  
Best Local Similarity 70.0%; Pred. No. 4.8e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12  
|| :|||||  
Db 581 CPPAEPRLPLE 590

RESULT 39  
US-08-738-975-5  
; Sequence 5, Application US/08738975  
; Patent No. 5880267  
; GENERAL INFORMATION:  
; APPLICANT: Fukuda, Michiko N.  
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
; TITLE OF INVENTION: Proteins  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/738,975  
; FILING DATE: herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/439,818  
; FILING DATE: 05-Dec-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 2251  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 778 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-738-975-5

Query Match 51.5%; Score 35; DB 2; Length 778;  
Best Local Similarity 70.0%; Pred. No. 4.8e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12  
|| :|||||  
Db 581 CPPAEPRLPLE 590

RESULT 40  
US-08-728-626-5  
; Sequence 5, Application US/08728626  
; Patent No. 5910451



GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
TITLE OF INVENTION: Trophinin and Trophinin-Assisting  
PROTEINS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICANT NUMBER: US/08/728,626  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: US 08/439,818  
FILING DATE: 12-MAY-1995  
APPLICATION NUMBER: US 08/317,522  
FILING DATE: 04-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 1563  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 778 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-728-626-5

Query Match 51.5%; Score 35; DB 2; Length 778;  
Best Local Similarity 70.0%; Pred. No. 4.8e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 1;

QY 5 CP--DPRPLE 12  
|| :|||||  
Db 581 CPPAEPRLPE 590

RESULT 41  
US-08-808-599A-5  
Sequence 5, Application US/08808599A  
Patent No. 6111089  
GENERAL INFORMATION:  
APPLICANT: Fukuda, Michiko N.  
TITLE OF INVENTION: Trophinin, Trophinin-Assisting  
PROTEINS AND METHODS TO INHIBIT IMPLANTATION  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/808,599A  
FILING DATE: 28-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,522  
FILING DATE: 04-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/439,818  
FILING DATE: 12-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 2256  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 778 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-808-599A-5

Query Match 51.5%; Score 35; DB 4; Length 778;  
Best Local Similarity 70.0%; Pred. No. 4.8e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12  
|| :|||||  
Db 581 CPPAEPRLPE 590

RESULT 42  
US-08-804-227C-2  
Sequence 2, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kunstock, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4472 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein

US-08-804-227C-2

Query Match 51.5%; Score 35; DB 2; Length 4472;  
Best Local Similarity 66.7%; Pred. No. 2.7e+03;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCDPRP 10  
| | | | |  
Db 3638 GTLVDPDP 3646

RESULT 43

US-08-465-794-14  
; Sequence 14, Application US/08465794  
; Patent No. 5886141

GENERAL INFORMATION:  
; APPLICANT: FOLKMAN, MOSES J.  
; APPLICANT: SHING, YUEN  
; APPLICANT: IGARASHI, KOICHI  
; TITLE OF INVENTION: SMOOTH MUSCLE MITOGEN AND ISOLATED DNA  
; TITLE OF INVENTION: CODING THEREFORE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
; ADDRESSEE: CUSHMAN  
; STREET: 130 WATER STREET  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: US  
; ZIP: 02109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,794  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/007,126  
; FILING DATE: 21-JAN-1993  
; APPLICATION NUMBER: US 07/994,776  
; FILING DATE: 22-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/872,597  
; FILING DATE: 23-APR-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/872,792  
; FILING DATE: 23-APR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/833,552  
; FILING DATE: 10-FEB-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/832,939  
; FILING DATE: 10-FEB-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/766,354  
; FILING DATE: 26-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/604,778  
; FILING DATE: 26-OCT-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RESNICK, DAVID S.  
; REGISTRATION NUMBER: 34235  
; REFERENCE/DOCKET NUMBER: 40435-CIP-8  
; TELEPHONE: (617) 523-3400  
; TELEFAX: (617) 523-6440  
; TELEX: 200291 STRE UR  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 60 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-465-794-14

Query Match 50.0%; Score 34; DB 2; Length 60;  
Best Local Similarity 85.7%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLLCDP 8  
| | | | |  
Db 13 GLLCDP 19

RESULT 44

US-09-049-813-14  
; Sequence 14, Application US/09049813  
; Patent No. 6013762

GENERAL INFORMATION:  
; APPLICANT: FOLKMAN, MOSES J.  
; APPLICANT: SHING, YUEN  
; APPLICANT: IGARASHI, KOICHI  
; TITLE OF INVENTION: SMOOTH MUSCLE MITOGEN AND ISOLATED DNA  
; TITLE OF INVENTION: CODING THEREFORE  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
; ADDRESSEE: CUSHMAN  
; STREET: 130 WATER STREET  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: US  
; ZIP: 02109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/049,813  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/465,794  
; FILING DATE:  
; APPLICATION NUMBER: US 08/007,126  
; FILING DATE: 21-JAN-1993  
; APPLICATION NUMBER: US 07/994,776  
; FILING DATE: 22-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/872,597  
; FILING DATE: 23-APR-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/872,792  
; FILING DATE: 23-APR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/833,552  
; FILING DATE: 10-FEB-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/832,939  
; FILING DATE: 10-FEB-1992  
; APPLICATION DATA:  
; APPLICATION NUMBER: US 07/766,354  
; FILING DATE: 26-SEP-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/604,778  
; FILING DATE: 26-OCT-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RESNICK, DAVID S.  
; REGISTRATION NUMBER: 34235  
; REFERENCE/DOCKET NUMBER: 40435-CIP-8  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 523-3400  
TELEFAX: (617) 523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-09-049-813-14

Query Match 50.0%; Score 34; DB 3; Length 60;  
Best Local Similarity 85.7%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLLCPDP 8  
Db 13 GLLCGDP 19

RESULT 45  
US-08-663-191A-1  
; Sequence 1, Application US/08663191A  
; Patent No. 6183971  
; GENERAL INFORMATION:  
; APPLICANT: Reiko SASADA, et al.  
; TITLE OF INVENTION: ANTIBODY, HYBRIDOMA AND USE THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wedgeroth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/663,191A  
; FILING DATE: 11-Jun-1996  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: <Unknown>  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lee Cheng  
; REGISTRATION NUMBER: 40,949  
; REFERENCE/DOCKET NUMBER: 96-0256/LC(WMC)/927  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-721-8200  
; TELEFAX: 202-721-8250  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80 amino acid residues  
; TYPE: amino acid  
; STRANDEDNESS: n/a  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-08-663-191A-1

Query Match 50.0%; Score 34; DB 4; Length 80;  
Best Local Similarity 85.7%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLLCPDP 8  
Db 12 GLLCGDP 18

Search completed: June 28, 2001, 11:56:11  
Job time: 192 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:57 ; Search time 42.81 Seconds  
(without alignments)  
21.352 Million cell updates/sec

Title: US-09-439-313-558  
Perfect score: 68  
Sequence: 1 AGLLCPDPRPLE 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	67.6	567	2 F64058	nitrate/nitrite se
2	44	64.7	397	2 H70753	hypothetical prote
3	44	64.7	757	2 T34362	hypothetical prote
4	41	60.3	362	1 C70646	probable pflA prot
5	41	60.3	397	2 B70815	probable transamin
6	41	60.3	430	2 B70813	hypothetical prote
7	41	60.3	568	2 JQ2206	UL46h protein - Ma
8	41	60.3	607	2 S01939	hypothetical prote
9	41	60.3	1255	1 A24571	protein-tyrosine k
10	40	58.8	140	2 G83562	hypothetical prote
11	40	58.8	210	2 G84499	hypothetical prote
12	40	58.8	261	2 D70871	hypothetical prote
13	40	58.8	276	2 G47301	VirB8 homolog - Bo
14	40	58.8	398	2 D85942	probable dehydrata
15	40	58.8	398	2 G65070	hypothetical prote
16	40	58.8	522	2 C81826	hypothetical integ
17	40	58.8	528	2 T52101	probable nuclear t
18	40	58.8	538	2 F86225	hypothetical prote
19	40	58.8	633	2 E75277	ABC transporter, A
20	39.5	58.1	399	2 D86177	protein FlpP19.3 l
21	39.5	58.1	595	2 A48878	uric acid/xanthine
22	39	57.4	146	2 I53030	submaxillary prote
23	39	57.4	432	2 I48343	interleukin-11 rec
24	39	57.4	527	2 T04329	importin alpha - t
25	39	57.4	532	2 T52102	probable nuclear t
26	39	57.4	532	2 T52268	importin alpha [va
27	39	57.4	535	2 T52098	probable nuclear t
28	39	57.4	1207	1 EGHU	epidermal growth f
29	38.5	56.6	529	2 T35966	probable secreted

30	38	55.9	338	2	G72241	electron transfer
31	38	55.9	350	2	G75571	MutT/nudix family
32	38	55.9	389	2	S74343	probable aspartate
33	38	55.9	475	2	T33943	hypothetical prote
34	38	55.9	483	2	S74637	lysine decarboxyla
35	38	55.9	578	2	E70645	hypothetical prote
36	38	55.9	999	1	S27756	receptor-like prot
37	38	55.9	1264	2	S64146	probable membrane
38	38	55.9	1321	2	JE0352	mucin MUC5B, trach
39	38	55.9	1690	2	T35694	ATP dependent DNA
40	38	55.9	2139	2	T18296	myosin heavy chain
41	37	54.4	121	2	B83432	conserved hypothet
42	37	54.4	246	2	S63639	hypothetical prote
43	37	54.4	269	2	H85408	GATA transcription
44	37	54.4	269	2	T05288	GATA-binding trans
45	37	54.4	415	2	S28088	gene B protein - y
46	37	54.4	518	2	S61920	B-alpha pheromone-
47	37	54.4	629	2	T23538	hypothetical prote
48	37	54.4	649	2	T04606	protein kinase hom
49	37	54.4	653	2	G70683	probable lepA - My
50	37	54.4	659	2	S30893	nrtC protein - Syn
51	37	54.4	779	2	T31732	hypothetical prote
52	37	54.4	935	2	E96806	hypothetical prote
53	37	54.4	1290	2	A57190	ebnerin precursor
54	36	52.9	116	2	T34800	hypothetical prote
55	36	52.9	141	2	S67149	hypothetical prote
56	36	52.9	174	2	B82747	hypothetical prote
57	36	52.9	180	2	T12493	hypothetical prote
58	36	52.9	229	2	C71493	probable uracil DN
59	36	52.9	249	2	S76255	hypothetical prote
60	36	52.9	251	1	G71366	probable uridylyate
61	36	52.9	267	2	T36675	probable integral
62	36	52.9	291	1	JN0064	insulin-like growt
63	36	52.9	310	2	JN0490	28K lipase precurs
64	36	52.9	332	2	H83924	tryptophanyl-tRNA
65	36	52.9	352	2	I40177	maleylacetate redu

ALIGNMENTS

RESULT 1  
F64058  
nitrate/nitrite sensor protein (EC 2.7.3.-) [similarity] - Haemophilus influenzae (st  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 21-Jul-2000  
C;Accession: F64058  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630  
A;Accession: F64058  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-567 <TIGR>  
A;Cross-references: GB:U32713; GB:I42023; NID:gl573231; PIDN:AAC21933.1; PID:gl573233  
C;Genetics:  
A;Start codon: GTG  
C;Superfamily: nitrate/nitrite sensor protein narX  
C;Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase  
F;373/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 67.6%; Score 46; DB 2; Length 567;  
Best Local Similarity 66.7%; Pred. No. 6.2;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AGLLCPDPRPLE 12

||| ||||| ::

Db 326 AGLPCCDPRIMQ 337

RESULT 2  
H70753  
hypothetical protein Rv1264 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: H70753  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: H70753  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-397 <COL>  
A;Cross-references: GB:Z77137; GB:AL123456; PIDN:CAB00890.1; PID:e254882;  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv1264

Query Match 64.7%; Score 44; DB 2; Length 397;  
Best Local Similarity 77.8%; Pred. No. 9.3;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPRPL 11  
|:|||||  
Db 269 LVCPDAPL 277

## RESULT 3

H34362  
hypothetical protein T19D12.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34362  
R;Favella, A.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid T19D12.  
A;Reference number: Z21513  
A;Accession: T34362  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-757 <FAV>  
A;Cross-references: EMBL:U41263; PIDN:AAC24426.1; GSPDB:GN00020; CESP:T19D12.6  
A;Experimental source: strain Bristol N2; clone T19D12  
C;Genetics:  
A;Gene: CESP:T19D12.6  
A;Map position: 2  
A;Introns: 55/1; 168/2; 201/2; 253/3; 292/3; 317/1; 373/1; 463/2; 539/3; 623/2; 652/1; 7

Query Match 64.7%; Score 44; DB 2; Length 757;  
Best Local Similarity 58.3%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLLCPDRPLE 12  
:|:|||||  
Db 25 SGVSCPDRPLQ 36

## RESULT 4

C70646  
Probable pflA protein - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: C70646  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: C70646  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-362 <COL>  
A;Cross-references: GB:Z83867; GB:AL123456; PIDN:CAB06292.1; PID:g32616  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: pflA  
C;Superfamily: conserved hypothetical protein MJ0808

Query Match 60.3%; Score 41; DB 1; Length 362;  
Best Local Similarity 66.7%; Pred. No. 26;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGLLCPDPR 9  
||:||||  
Db 162 AGYMCPEPR 170

## RESULT 5

B70815  
probable transaminase (EC 2.6.1.1-) Rv0858c [similarity] - Mycobacterium tuberculosis  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: B70815  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: B70815  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-397 <COL>  
A;Cross-references: GB:AL022004; GB:AL123456; PIDN:CAA17664.1; PID:g291  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: Rv0858c  
C;Superfamily: aspartate transaminase  
C;Keywords: aminotransferase

Query Match 60.3%; Score 41; DB 2; Length 397;  
Best Local Similarity 87.5%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11  
||:||||  
Db 318 LCADPRPL 325

## RESULT 6

B70813  
hypothetical protein Rv0842 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: B70813  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: B70813

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-430 <COL>  
A;Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17648.1; PID:e125398  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: RV0842

Query Match 60.3%; Score 41; DB 2; Length 430;  
Best Local Similarity 63.6%; Pred. No. 31;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPL 11  
Db 120 AGLVCDAPRPV 130  
|||:| |||:

RESULT 7  
JQ2206  
UL46h protein - Marek's disease virus  
C;Species: Marek's disease virus  
C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 08-Oct-1999  
C;Accession: JQ2206  
R;Yanagida, N.; Yoshida, S.; Nazerian, K.; Lee, L.F.  
J. Gen. Virol. 74, 1837-1845, 1993  
A;Title: Nucleotide and predicted amino acid sequences of Marek's disease virus homologues  
A;Reference number: JQ2199; MUID:93389438  
A;Accession: JQ2206  
A;Molecule type: DNA  
A;Residues: 1-568 <YAN>  
A;Cross-references: GB:L10283; NID:g388703; PIDN:AAA03151.1; PID:g388709  
A;Experimental source: strain GA

Query Match 60.3%; Score 41; DB 2; Length 568;  
Best Local Similarity 54.5%; Pred. No. 40;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRPLE 12  
Db 479 GVLCPAPQPAD 489  
|:|:| |:|:

RESULT 8  
S01939  
hypothetical protein 2 - beet western yellows virus (isolate FL1)  
C;Species: beet western yellows virus  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Sep-1999  
C;Accession: S01939  
R;Veidt, I.; Lot, H.; Leiser, M.; Scheidecker, D.; Guilley, H.; Richards, K.; Jonard, G.  
Nucleic Acids Res. 16, 9917-9932, 1988  
A;Title: Nucleotide sequence of beet western yellows virus RNA.  
A;Reference number: S01935; MUID:89057523  
A;Accession: S01939  
A;Molecule type: genomic RNA  
A;Residues: 1-607 <VEI>  
A;Cross-references: EMBL:X13063; NID:g62294; PIDN:CAA31463.1; PID:g62296  
C;Superfamily: potato leaf roll virus 70K protein

Query Match 60.3%; Score 41; DB 2; Length 607;  
Best Local Similarity 85.7%; Pred. No. 43;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 CPDPRPL 11  
Db 62 CPEPRPL 68  
||:||||

RESULT 9  
A24571  
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e  
C;Species: Homo sapiens (man)  
C;Date: 25-Oct-1987 #sequence\_revision 06-Dec-1996 #text\_change 11-Jun-1999  
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622  
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T  
Nature 319, 230-234, 1986  
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor  
A;Reference number: A24571; MUID:86118663  
A;Accession: A24571  
A;Molecule type: mRNA  
A;Residues: 1-1255 <YAM>  
A;Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198  
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985  
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid  
A;Reference number: A25491; MUID:86016729  
A;Accession: A25491  
A;Molecule type: DNA  
A;Residues: 737-1031 <SEM>  
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282  
R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg  
Science 230, 1132-1139, 1985  
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal  
A;Reference number: A44188; MUID:86070181  
A;Accession: A44188  
A;Molecule type: DNA  
A;Residues: 740-910 <COU1>  
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989  
A;Accession: B44188  
A;Molecule type: mRNA  
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>  
A;Cross-references: GB:M11730; NID:g183986  
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985  
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A;Reference number: I59509; MUID:85272597  
A;Accession: I59509  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 832-909 <REX>  
A;Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808  
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987  
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional  
A;Reference number: I57622; MUID:87286898  
A;Accession: I57622  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-191 <TAL>  
A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332  
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30% of human breast cancers.  
C;Genetics:  
A;Gene: GDB:ERBB2; NGL; NEU; HER-2  
A;Cross-references: GDB:L20613; OMIM:164870  
A;Map position: 17q21.1-17q21.1  
A;Introns: 25/1; 75/3; 147/1; 883/3  
A;Note: the list of introns is incomplete  
C;Function:  
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
C;Superfamily: epidermal growth factor receptor; protein kinase homology  
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
F;70-304/Domain: extracellular #status predicted <EXT>  
F;395-605/Domain: EGF receptor extracellular domain repeat <EE1>  
F;654-675/Domain: EGF receptor extracellular domain repeat <TMM>  
F;676-1255/Domain: transmembrane #status predicted <INT>  
F;718-983/Domain: intracellular #status predicted <KIN>  
F;726-734/Region: protein kinase ATP-binding motif  
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F:753/Active site: Lys #status predicted  
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 60.3%; Score 41; DB 1; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 GLLCPDPRP 10  
DB 1029 GFFCPDPAP 1037

RESULT 10  
G83562  
hypothetical protein PA0664 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83562  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, L.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83562  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-140 <STO>  
A:Cross-references: GB:AE004501; GB:AE004091; NID:99946537; PIDN:AAG04053.1; GSPDB:GN00139  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0664

Query Match 58.8%; Score 40; DB 2; Length 140;  
Best Local Similarity 60.0%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 2; Indels 4; Gaps 1;  
QY 2 GLLCP---DPRPLE 12  
DB 117 GRLCPMTEASPRPLE 131

RESULT 11  
G84499  
hypothetical protein At2g11920 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: G84499  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: G84499  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-210 <STO>  
A:Cross-references: GB:AE002093; NID:94557064; PIDN:AAD22503.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g11920  
A:Map position: 2

Query Match 58.8%; Score 40; DB 2; Length 210;  
Best Local Similarity 66.7%; Pred. No. 22;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 3 LLLCPDPRPL 11  
DB 169 LVCPDPRSI 177

RESULT 12

D70871  
hypothetical protein Rv1457c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D70871  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: D70871  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-261 <COL>  
A:Cross-references: GB:AL021184; GB:AL123456; NID:g3261498; PIDN:CAAL5987.1; PID:g279  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv1457c  
C:Superfamily: Streptomyces peucetius daunorubicin resistance protein

Query Match 58.8%; Score 40; DB 2; Length 261;  
Best Local Similarity 70.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AGLLCPDPRP 10  
DB 11 AGTFSPDPRP 20

RESULT 13

G47301  
VirB8 homolog - Bordetella pertussis  
C:Species: Bordetella pertussis  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
C:Accession: G47301  
R:Weiss, A.A.; Johnson, F.D.; Burns, D.L.  
Proc. Natl. Acad. Sci. U.S.A. 90, 2970-2974, 1993  
A:Title: Molecular characterization of an operon required for pertussis toxin secreti  
A:Reference number: A47301; MUID:93219406  
A:Contents: BP338  
A:Accession: G47301  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-276 <WEI>  
A:Note: sequence extracted from NCBI backbone (NCBIN:128775, NCBIP:128779)

Query Match 58.8%; Score 40; DB 2; Length 276;  
Best Local Similarity 63.6%; Pred. No. 29;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AGLLCPDPRPL 11  
DB 40 ARIVMPDPRPL 50

RESULT 14

D85942  
probable dehydratase ygeX [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: D85942  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551





C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C;Accession: E75277  
R;White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896  
A;Accession: E75277  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-633 <WHI>  
A;Cross-references: GB:AE002071; GB:AE000513; NID:g6460218; PIDN:AAF11948.1; PID:g646022  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR2404  
A;Map position: 1  
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 58.8%; Score 40; DB 2; Length 633;  
Best Local Similarity 87.5%; Pred. NO. 65;  
Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

OY 4 LCPDRPL 11  
I | | | | |  
Db 369 LAPDRPL 376

RESULT 20  
D86177  
Protein F19p19.3 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: D86177  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719  
A;Accession: D86177  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-399 <STO>  
A;Cross-references: GB:AE005172; NID:g2341026; PIDN:AAE70424.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F19p19.3  
A;Map position: 1

Query Match 58.1%; Score 39.5; DB 2; Length 399;  
Best Local Similarity 63.6%; Pred. NO. 50;  
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

OY 2 GLICPDRPL 11  
I | | | | |  
Db 252 GLFCDPPIKPI 262

RESULT 21  
A48878  
uric acid/xanthine transport protein - Emericella nidulans  
N;Alternate names: uric acid/xanthine permease  
C;Species: Emericella nidulans, Aspergillus nidulans  
C;Date: 19-May-1995 #sequence\_revision 19-May-1995 #text\_change 04-Sep-1998  
C;Accession: A48878; S36031  
R;Gorfinkiel, L.; Dhalluin, G.; Scazzocchio, C.

J. Biol. Chem. 268, 23376-23381, 1993  
A;Title: Sequence and regulation of the uapA gene encoding a uric acid-xanthine perme  
A;Reference number: A48878; MUID:94043131  
A;Accession: A48878  
A;Molecule type: DNA  
A;Residues: 1-595 <GOR>  
A;Cross-references: EMBL:X71807; NID:g298063; PID:g298064  
C;Genetics:  
A;Gene: uapA  
A;Introns: 161/2; 343/3; 554/2  
A;Superfamily: uric acid/xanthine transport protein  
C;Keywords: transmembrane protein

Query Match 58.1%; Score 39.5; DB 2; Length 595;  
Best Local Similarity 61.5%; Pred. NO. 74;  
Matches 8; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

OY 2 GLICPD---PRPL 11  
I | | | | |  
Db 273 GMLCPSATAPRPL 285

RESULT 22  
I53030  
submaxillary protein SMR1 precursor - black rat  
C;Species: Rattus rattus (black rat, roof rat)  
A;Note: this sequence is derived from an apparently genuine specimen of this problema  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Jun-2000  
C;Accession: I53030  
R;Singer, M.; Courty, Y.; Rougeon, F.  
DNA Cell Biol. 14, 137-144, 1995  
A;Title: Recent evolution of genes encoding the prohormone-like protein SMR1 in the r  
A;Reference number: I53030; MUID:95169272  
C;Accession: I53030  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-146 <RES>  
A;Cross-references: EMBL:X77819; NID:g732921; PIDN:CAA54834.1; PID:g732922  
C;Genetics:  
A;Gene: VCS-alpha  
C;Superfamily: proline-rich peptide P-B  
C;Keywords: glycoprotein  
F;1-18/Domains: signal sequence #status predicted <SIG>  
F;19-146/Product: submaxillary protein SMR1 #status predicted <MAT>  
F;129,136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 57.4%; Score 39; DB 2; Length 146;  
Best Local Similarity 77.8%; Pred. NO. 23;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 LLCPPDRPL 11  
I | | | | |  
Db 98 LTAPDRPL 106

RESULT 23  
I48343  
interleukin-11 receptor alpha-chain precursor - mouse  
N;Alternate names: cytokine type 1 receptor  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 31-Jan-2000  
C;Accession: I48343; S51619  
R;Neuhaus, H.; Bettenhausen, B.; Bilinski, P.; Simon-Chazottes, D.; Guenet, J.L.; Gos  
Dev Biol. 166, 521-542, 1994  
A;Title: etl2, a novel putative type-1 cytokine receptor expressed during mouse embryo  
A;Reference number: I48343  
A;Accession: I48343  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-432 <RES>  
A;Cross-references: EMBL:X74953; NID:g673437; PIDN:CAA52908.1; PID:g673438

R;Hilton, D.J.; Hilton, A.A.; Raicevic, A.; Rakar, S.; Harrison-Smith, M.; Gough, N.M.; EMBO J. 13, 4765-4775, 1994  
A;Title: Cloning of a murine IL-11 receptor alpha-chain; requirement for gp130 for high  
A;Reference number: S51619; MUID:95045367  
A;Accession: S51619  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-432 <HIL>  
A;Cross-references: EMBL:U14412; NID:g576454; PIDN:AAA53248.1; PID:g576455  
C;Genetics:  
A;Gene: Et12/IL11  
C;Superfamily: ciliary neurotrophic factor receptor; cytokine receptor homology; immunog  
C;Keywords: cytokine receptor  
F;120-310/Domain: cytokine receptor homology <CRS>  
  
Query Match 57.4%; Score 39; DB 2; Length 432;  
Best Local Similarity 77.8%; Pred. No. 65;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 LCPDPRPLE 12  
| | | | | | | |  
Db 352 LQDPRPLD 360  
  
RESULT 24  
T04329  
importin alpha - tomato  
N;Alternate names: KAP alpha protein  
C;Species: Lycopersicon esculentum (tomato)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 29-Sep-1999  
C;Accession: T04329  
R;Kunik, T.; Mizrachy, L.; Citovsky, V.; Gafni, Y.  
submitted to the EMBL Data Library, June 1998  
A;Reference number: Z15286  
A;Accession: T04329  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-527 <KUN>  
A;Cross-references: EMBL:AF017252; NID:g3228369; PIDN:AAC23722.1; PID:g3228370  
C;Genetics:  
A;Gene: KAP  
C;Superfamily: pendulin  
  
Query Match 57.4%; Score 39; DB 2; Length 527;  
Best Local Similarity 85.7%; Pred. No. 79;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 LLCPPDR 9  
|:|:|:|:|:|  
Db 422 LVCPDPR 428  
  
RESULT 25  
T52102  
probable nuclear transport factor importin alpha-like protein [imported] - Arabidopsis t  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 03-Nov-2000  
C;Accession: T52102  
R;Merkle, T.  
submitted to the EMBL Data Library, March 1998  
A;Reference number: Z25955  
A;Accession: T52102  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-532 <MER>  
A;Cross-references: EMBL:Y15224; PIDN:CAA75513.1  
A;Experimental source: ecotype Columbia; vegetative; 3 weeks old  
C;Genetics:  
A;Gene: Impal  
C;Superfamily: pendulin

Query Match 57.4%; Score 39; DB 2; Length 532;  
Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 LLCPPDR 9  
|:|:|:|:|:|  
Db 421 LVCPDPR 427  
  
RESULT 26  
T52268  
importin alpha [validated] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 03-Nov-2000  
C;Accession: T52268  
R;Smith, H.M.; Hicks, G.R.; Raikhel, N.V.  
Plant Physiol. 114, 411-417, 1997  
A;Title: Importin alpha from Arabidopsis thaliana is a nuclear import receptor that r  
A;Reference number: Z26010; MUID:97336305  
A;Accession: T52268  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-532 <SMI>  
A;Cross-references: EMBL:AF077528; PIDN:AAC27644.1  
A;Experimental source: cultivar Columbia  
C;Function:  
A;Description: involved in protein import in nucleus; binds to nuclear localization s  
C;Superfamily: pendulin  
  
Query Match 57.4%; Score 39; DB 2; Length 532;  
Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 LLCPPDR 9  
|:|:|:|:|:|  
Db 421 LVCPDPR 427  
  
RESULT 27  
T52098  
probable nuclear transport factor importin alpha [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 03-Nov-2000  
C;Accession: T52098  
R;Schledz, M.; Leclerc, D.; Neuhaus, G.; Merkle, T.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z25951  
A;Accession: T52098  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-535 <SCH>  
A;Cross-references: EMBL:Y14615; PIDN:CAA74965.1  
A;Experimental source: ecotype Columbia; vegetative tissue; 3 weeks old  
C;Genetics:  
A;Gene: Impa-2  
C;Superfamily: pendulin  
  
Query Match 57.4%; Score 39; DB 2; Length 535;  
Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 LLCPPDR 9  
|:|:|:|:|:|  
Db 426 LVCPDPR 432  
  
RESULT 28  
EGHU  
epidermal growth factor precursor [validated] - human  
N;Alternate names: urogastrome precursor

C:Species: Homo sapiens (man)  
C:Date: 30-Nov-1980 #sequence revision 14-Aug-1998 #text\_change 08-Dec-2000  
C:Accession: A25531; A01388; A33517; A29721; S45282; S45283  
R:Bell, G.I.; Fong, N.M.; Stempien, M.M.; Wormsted, M.A.; Caput, D.; Ku, L.; Urdea, M.S.  
Nucleic Acids Res. 14, 8427-8446, 1986  
A:Title: Human epidermal growth factor precursor: cDNA sequence, expression in vitro and  
A:Reference number: A25531; MUID:87066721  
A:Accession: A25531  
A:Molecule type: mRNA  
A:Residues: 1-1207 <BEL>  
A:Cross-references: EMBL:X04571; NID:g311120; PIDN:CAA28240.1; PID:g311121  
A:Note: 708-Met was also found  
R:Gregory, H.; Preston, B.M.  
Int. J. Pept. Protein Res. 9, 107-118, 1977  
A:Title: The primary structure of human urogastrone.  
A:Reference number: A01388; MUID:77117897  
A:Accession: A01388  
A:Molecule type: protein  
A:Residues: 971-1023 <GRE>  
A:Note: Some of the molecules lack Arg-1023  
R:Furuya, M.; Akashi, S.; Hirayama, K.  
Biochem. Biophys. Res. Commun. 163, 1100-1106, 1989  
A:Title: The primary structure of human EGF produced by genetic engineering, studied by  
A:Reference number: A33517; MUID:89391964  
A:Accession: A33517  
A:Molecule type: protein  
A:Residues: 971-1023 <FUR>  
R:Tsumoto, K.; Nakamura, H.; Sakamoto, S.  
Biochem. Biophys. Res. Commun. 145, 126-133, 1987  
A:Title: Purification and characterization of high molecular weight human epidermal growth  
A:Reference number: A29721; MUID:87241488  
A:Accession: A29721  
A:Molecule type: protein  
A:Residues: 829-834, 'X', 836-839, 'X', 841-845, 'X', 847-848 <TSU>  
A:Note: this is the amino-terminal sequence of a high molecular weight form of EGF, isolated  
R:Svoboda, M.; Bauhofer, A.; Schwind, P.; Bade, E.; Rasched, I.; Przybylski, M.  
Biochim. Biophys. Acta 1206, 35-41, 1994  
A:Title: Structural characterization and biological activity of recombinant human epidermal  
A:Reference number: S45282; MUID:94242778  
A:Accession: S45282  
A:Molecule type: protein  
A:Residues: 'M', 971-1023 <SVO>  
A:Note: expressed recombinant protein  
A:Accession: S45283  
A:Molecule type: protein  
A:Residues: 'MKKYP', 970-1023 <SV2>  
A:Note: expressed recombinant protein  
C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiation of  
gastrointestinal cell proliferation.  
C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in some  
C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound protein  
C:Genetics:  
A:Gene: GDB:EGF  
A:Cross-references: GDB:119105; OMIM:131530  
A:Map position: 4q25-4q25  
A:Introns: 43/1; 109/3; 170/2; 246/2; 314/1; 356/3; 397/1; 438/1; 480/1; 525/3; 575/2; 6  
C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-contain  
C:Keywords: duplication; growth factor; tandem repeat; transmembrane protein  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-1207/Product: epidermal growth factor propoetoin, membrane-bound form #status predi  
F:23-1032/Domain: extracellular #status predicted <EXT>  
F:43-479/Region: EGF precursor long repeat <LR1>  
F:46-85/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
F:86-127/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
F:128-169/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
F:170-211/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
F:212-256/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
F:257-301/Domain: LDL receptor YWTD-containing repeat homology <EG1>  
F:318-354/Domain: EGF homology <EG2>  
F:360-395/Domain: EGF homology <EG3>  
F:401-436/Domain: EGF homology <EG4>  
F:439-476/Domain: EGF homology <EG5>

F:480-954/Region: EGF precursor long repeat <LR2>  
F:483-523/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
F:524-566/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
F:567-609/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
F:610-653/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
F:654-694/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
F:695-737/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
F:745-780/Domain: EGF homology <EG5>  
F:835-868/Domain: EGF homology <EG6>  
F:874-910/Domain: EGF homology <EG7>  
F:916-951/Domain: EGF homology <EG8>  
F:971-1023/Product: epidermal growth factor #status experimental <EGF>  
F:976-1012/Domain: EGF homology <EG9>  
F:1033-1057/Domain: transmembrane #status predicted <TMN>  
F:1058-1207/Domain: intracellular #status predicted <INT>  
F:318-330, 325-339, 341-354, 360-371, 367-380, 382-395, 401-412, 408-421, 423-436, 439-451, 447  
fide bonds: #status predicted  
F:976-990, 984-1001, 1003-1012/Disulfide bonds: #status experimental

Query Match 57.4%; Score 39; DB 1; Length 1207;  
Best Local Similarity 66.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPDPRP 10  
||:||||  
Db 948 GLICPDSTP 956

RESULT 29  
T35966  
Probable secreted peptidase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T35966  
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M  
submitted to the EMBL Data Library, January 1999  
A:Reference number: 221551  
A:Accession: T35966  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-529 <SEE>  
A:Cross-references: EMBL:AL035161; PIDN:CAA22727.1; GSPDB:GN00070; SCOEDB:SC9C7.15C  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC9C7.15C

Query Match 56.6%; Score 38.5; DB 2; Length 529;  
Best Local Similarity 53.3%; Pred. No. 96;  
Matches 8; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 1 AGLIC--PDPRLP 12  
|:| ||:||||  
Db 494 ADAVCQALPDPKPLE 508

RESULT 30  
G72241  
Electron transfer flavoprotein, alpha subunit - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: G72241  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316  
A:Accession: G72241  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-338 <ARN>



A;Cross-references: GB:AE001800; GB:AE000512; NID:g4982090; PIDN:AAD36598.1; PID:g498209  
A;Experimental source: strain MSB8  
C;Genetics: TM1531  
C;Superfamily: electron transfer flavoprotein alpha chain

Query Match 55.9%; Score 38; DB 2; Length 338;  
Best Local Similarity 60.0%; Pred. No. 75;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLLCPDPRP 10  
| : : | | | | |  
Db 149 ATIMCPDHRP 158

RESULT 31  
G75571

MutF/nudix family protein - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C;Accession: G75571  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896  
A;Accession: G75571  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-350 <WHI>  
A;Cross-references: GB:AE001864; GB:AE000513; NID:g6457659; PIDN:AAF09597.1; PID:g645766  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR0004  
A;Map position: 1

Query Match 55.9%; Score 38; DB 2; Length 350;  
Best Local Similarity 70.0%; Pred. No. 77;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRPL 11  
| | | | | | |  
Db 72 GLRCPNLRPL 81

RESULT 32  
S74343

probable aspartate transaminase (EC 2.6.1.1) aspC sll0402 [similarity] - Synechocystis s  
N;Alternate names: hypothetical protein sll0402  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C;Accession: S74343  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.  
A;Reference number: S74322; MUID:97061201  
A;Accession: S74343  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-389 <KAN>  
A;Cross-references: EMBL:D64001; GB:AB001339; NID:g1001102; PIDN:BAA10261.1; PID:g100112  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Genetics:  
A;Gene: aspC  
C;Superfamily: aspartate transaminase  
C;Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate  
F:238/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 55.9%; Score 38; DB 2; Length 389;  
Best Local Similarity 66.7%; Pred. No. 86;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGLLCPDPR 9  
| | | | | | |  
Db 312 AGLSCPMPK 320

RESULT 33  
T33943

hypothetical protein C01B4.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T33943  
R;Smith, A.; Wamsley, P.; Fronick, W.  
submitted to the EMBL Data Library, February 1999  
A;Description: The sequence of C. elegans cosmid C01B4.  
A;Reference number: Z21443  
A;Accession: T33943  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-475 <SMI>  
A;Cross-references: EMBL:AF125952; PIDN:AAD14699.1; GSPDB:GN00023; CESP:C01B4.7  
A;Experimental source: strain Bristol N2; clone C01B4  
C;Genetics:  
A;Gene: CESP:C01B4.7  
A;Map position: 5  
A;Introns: 45/2; 80/1; 118/2; 189/3; 239/2; 340/3; 433/3

Query Match 55.9%; Score 38; DB 2; Length 475;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10  
| : | | | | |  
Db 423 GILCPDNTTP 431

RESULT 34  
S74637

lysine decarboxylase cad - Synechocystis sp. (strain PCC 6803)  
N;Alternate names: hypothetical protein sll1683  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C;Accession: S74637  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
s.  
A;Reference number: S74322; MUID:97061201  
A;Accession: S74637  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-483 <KAN>  
A;Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAA16789.1; PID:g165  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Genetics:  
A;Gene: cad  
C;Superfamily: lysine decarboxylase cad

Query Match 55.9%; Score 38; DB 2; Length 483;  
Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGLLCPDP 8  
| | | | | | |

Db 431 AGLICPYP 438

RESULT 35

E70645

hypothetical protein RV3132C - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: E70645

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: E70645

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-578 <COL>

A:Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CAB06282.1; PID:e290953;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3132C

Query Match 55.9%; Score 38; DB 2; Length 578;

Best Local Similarity 70.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPDRPL 11

Db 122 GLIEDPKPL 131

RESULT 36

S27756

receptor-like protein kinase 5 (EC 2.7.1.1-) precursor - Arabidopsis thaliana

N:Alternate names: protein F2009.180

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 11-Jun-1999

C:Accession: S27756; T04620

R:Walker, J.C.

submitted to the EMBL Data Library, February 1992

A:Description: Receptor-like protein kinase genes of Arabidopsis thaliana.

A:Reference number: S27754

A:Accession: S27756

A:Molecule type: mRNA

A:Residues: 1-999 <WAL>

A:Cross-references: EMBL:M84660; NID:gl66849; PIDN:AAA32859.1; PID:gl66850

A:Experimental source: strain Columbia

R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X

submitted to the Protein Sequence Database, October 1998

A:Reference number: Z15380

A:Accession: T04620

A:Molecule type: DNA

A:Residues: 1-999 <BEV>

A:Cross-references: EMBL:AL021749

A:Experimental source: cultivar Columbia; BAC clone F2009

C:Genetics:

A:Gene: RLK5

A:Map position: 4

A:Introns: 868/1

A>Note: F2009.180

C:Superfamily: protein kinase xa21; leucine-rich alpha-2-glycoprotein repeat homology; F

C:Keywords: Arp; autophosphorylation; glycoprotein; magnesium; phosphotransferase; recep

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-999/Product: receptor-like protein kinase #status predicted <MAT>

F:66-89/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>

F:90-114/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>

F:115-139/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:140-163/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

F:164-187/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>

F:188-212/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>

F:213-236/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>

F:237-260/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>

F:261-284/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>

F:285-307/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>

F:308-331/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>

F:332-355/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>

F:356-379/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>

F:380-403/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>

F:404-427/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>

F:428-451/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR16>

F:452-475/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR17>

F:476-499/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR18>

F:500-523/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR19>

F:524-547/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR20>

F:548-570/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR21>

F:571-593/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR22>

F:681-971/Domain: protein kinase Arp-binding motif

F:689-697/Region: protein kinase Arp-binding site: carbohydrate (Asn) (covalent) #sta

F:98,102,150,185,210,269,282,452,576/Binding site: Lys, Glu, Asp, Lys #status predicted

F:711,737,819,821/Active site: Lys, Glu, Asp, Lys #status predicted

F:824,828/Binding site: magnesium (Asn, Asp) #status predicted

Query Match 55.9%; Score 38; DB 1; Length 999;

Best Local Similarity 70.0%; Pred. No. 2.1e+02;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLICPDRPL 11

Db 943 GLICTSPLPL 952

RESULT 37

S64146

probable membrane protein YGL133w - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein G2842

C:Species: Saccharomyces cerevisiae

C>Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 17-Apr-1998

C:Accession: S64146; S71739

R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64144

A:Accession: S64146

A:Molecule type: DNA

A:Residues: 1-1264 <ESC>

A:Cross-references: EMBL:Z72655; NID:gl322701; PID:e243361; PID:gl322702; MIPS:YGL133

A:Experimental source: strain S288C

R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.

Yeast 12, 887-892, 1996

A:Title: Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces cerevisiae chro

A:Reference number: S71733; MUID:96437978

A:Accession: S71739

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1264 <ESW>

A:Cross-references: EMBL:X92670

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Map position: 7L

C:Keywords: transmembrane protein

F:549-565/Domain: transmembrane #status predicted <TMM>

Query Match 55.9%; Score 38; DB 2; Length 1264;

Best Local Similarity 66.7%; Pred. No. 2.7e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPDRPL 11

Db 9 ILLDPKPL 17

```
RESULT 38
JE0352
mucin MUC5B, tracheobronchial - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C;Accession: JE0352
R;Offner, G.D.; Nunes, D.P.; Keates, A.C.; Afdhal, N.H.; Troxler, R.F.
Biochem. Biophys. Res. Commun. 251, 350-355, 1998
A;Title: The amino-terminal sequence of MUC5B contains conserved multifunctional D domain
A;Reference number: JE0352; MUID:99009274
A;Accession: JE0352
A;Molecule type: mRNA
A;Residues: 1-1321 <OFF>
A;Cross-references: GB:AF086604; NID:g3789926; PIDN:AAC67545.1; PID:g3789927
C;Comment: This protein is large multimeric glycoproteins which is secreted by epithelial
C;Genetics:
A;Gene: MUC5B

Query Match 55.9%; Score 38; DB 2; Length 1321;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CPDPRPL 11
      ||||| ||
Db 244 CPDPLPL 250

RESULT 39
T35694
ATP dependent DNA helicase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35694
R;Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A;Reference number: 221587
A;Accession: T35694
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1690 <HAR>
A;Cross-references: EMBL:AL031031; PIDN:CAA19862.1; GSPDB:GN00070; SCOEDB:SC7C7.16c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC7C7.16c

Query Match 55.9%; Score 38; DB 2; Length 1690;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPLE 12
      | | | | | | | |
Db 1204 APLLLPPHPPLE 1215

RESULT 40
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: 218865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: EMBL:I03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C;Genetics:
A;Gene: mhca
```

```
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Query Match 55.9%; Score 38; DB 2; Length 2139;
Best Local Similarity 66.7%; Pred. No. 4.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLICPDPRP 10
      | : | | : | | |
Db 24 GILIPNRP 32

RESULT 41
B83432
conserved hypothetical protein in type III secretion PA1701 [imported] - Pseudomonas
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83432
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: B83432
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-121 <STO>
A;Cross-references: GB:AE004597; GB:AE004091; NID:g9947671; PIDN:AAG05090.1; GSPDB:GN
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA1701
C;Superfamily: Yersinia pestis hypothetical protein Y0045

Query Match 54.4%; Score 37; DB 2; Length 121;
Best Local Similarity 70.0%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPDPRPLE 12
      | | | | | | | |
Db 31 LLAPDGRPVE 40

RESULT 42
S63639
hypothetical protein 246 - Allomyces macrogynus mitochondrion
C;Species: mitochondrion Allomyces macrogynus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
C;Accession: S63639
R;Paquin, B.; Lang, B.F.
J. Mol. Biol. 255, 688-701, 1996
A;Title: The mitochondrial DNA of Allomyces macrogynus: the complete genomic sequence
A;Reference number: S63635; MUID:96226032
A;Accession: S63639
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-246 <PAQ>
A;Cross-references: EMBL:U41288
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C;Genetics:
A;Genome: mitochondrion
A;Introns: 243/3
C;Superfamily: probable RNA maturase
C;Keywords: mitochondrion

Query Match 54.4%; Score 37; DB 2; Length 246;
Best Local Similarity 75.0%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11
```

Db 99 LAPDPKPL 106  
| |||: ||  
RESULT 43  
H85408  
GATA transcription factor 3 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Mar-2001  
C:Accession: H85408  
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: H85408  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-269 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7270419; PIDN:CAB80185.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g34680  
A:Map position: 4  
C:Superfamily: Arabidopsis thaliana GATA transcription factor 4

Query Match 54.4%; Score 37; DB 2; Length 269;  
Best Local Similarity 50.0%; Pred. No. 87;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLLCDPRPLE 12  
: | | | : | | :  
Db 213 SGRLCPEYRPAD 224

RESULT 44  
T05288  
GATA-binding transcription factor homolog 3 [imported] - Arabidopsis thaliana  
N:Alternate names: protein T4L20.260  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 31-Dec-2000  
C:Accession: T05288; T52105  
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De  
ewes, H.W.; Mayer, K.F.X.; Schueller, C.  
submitted to the Protein Sequence Database, September 1998  
A:Reference number: Z15406  
A:Accession: T05288  
A:Molecule type: DNA  
A:Residues: 1-269 <BEV>  
A:Cross-references: EMBL:AL023094  
A:Experimental source: cultivar Columbia; BAC clone T4L20  
A:Genetics: B1  
R:Teakle, G.R.; Gilmartin, P.M.  
submitted to the EMBL Data Library, June 1997  
A:Description: Two types of GATA factor are found in fungi but are uniquely represented  
A:Reference number: Z25956  
A:Accession: T52105  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-54, 'Q', 56-269 <TEA>  
A:Cross-references: EMBL:Y13650; PIDN:CAA74001.1  
A:Experimental source: ecotype Columbia  
A:Genetics: B2  
C:Genetics: <B1>  
A:Gene: T4L20.260  
A:Map position: 4  
A:Introns: 94/3  
A:Note: T4L20.260  
C:Genetics: <B2>  
C:Superfamily: Arabidopsis thaliana GATA transcription factor 4

Best Local Similarity 50.0%; Pred. No. 87;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLLCDPRPLE 12  
: | | | : | | :  
Db 213 SGRLCPEYRPAD 224

RESULT 45  
S28088  
gene B protein - yeast (Kluyveromyces marxianus var. drosophilum) plasmid PKD1  
C:Species: Kluyveromyces marxianus var. drosophilum  
C:Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 29-Oct-1999  
C:Accession: S28088  
R:Chen, X.J.; Saliola, M.; Falcione, C.; Bianchi, M.M.; Fukuhara, H.  
Nucleic Acids Res. 14, 4471-4481, 1986  
A:Title: Sequence organization of the circular plasmid PKD1 from the yeast Kluyveromy  
A:Reference number: S28086; MUID:86232585  
A:Accession: S28088  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-415 <CHE>  
A:Cross-references: EMBL:X03961; NID:g2800; PIDN:CAA27593.1; PID:g2803  
C:Genetics:  
A:Gene: B  
A:Genome: plasmid

Query Match 54.4%; Score 37; DB 2; Length 415;  
Best Local Similarity 70.0%; Pred. No. 1-3e+02;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 LLCPPDRPLE 12  
| | | | | | | |  
Db 132 LLCPSPELLE 141

Search completed: June 28, 2001, 11:56:59  
Job time: 240 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:22 ; Search time 22.61 seconds  
(without alignments)  
18.181 Million cell updates/sec

Title: US-09-439-313-558  
Perfect score: 68  
Sequence: 1 AGLICPDRPLE 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	46	67.6	567	1	NARQ_HAEIN	P44604 haemophilus
2	44	64.7	397	1	YC64_MYCTU	Q11055 mycobacteri
3	41	60.3	607	1	V66K_BWYVF	P09506 beet wester
4	41	60.3	1255	1	ERB2_HUMAN	P04626 homo sapien
5	40	58.8	398	1	DPAL_ECOLI	Q46804 escherichia
6	39.5	58.1	615	1	UAPA_EMENI	Q07307 emericella
7	39	57.4	527	1	IMA_LYCES	O22478 lycopersico
8	39	57.4	596	1	IMAL_ARATH	Q96321 arabidopsis
9	39	57.4	1207	1	EGF_HUMAN	P01133 homo sapien
10	39	57.4	1687	1	Z142_HUMAN	P52746 homo sapien
11	38	55.9	338	1	ETFA_MEGEL	O85692 megasphaera
12	38	55.9	389	1	AAT_SYNY3	Q55128 synechocyst
13	38	55.9	395	1	DIHR_MANSE	P35464 manduca sex
14	38	55.9	999	1	RLK5_ARATH	P47735 arabidopsis
15	38	55.9	1264	1	YGN3_YEAST	P53125 saccharomyc
16	37	54.4	391	1	P53_CAVPO	Q9wur6 cavia porce
17	37	54.4	415	1	REPL_KLULA	P13775 kluyveromyc
18	37	54.4	518	1	BAR2_SCHCO	Q05659 schizophyll
19	37	54.4	567	1	GPV_MOUSE	O08742 mus musculu
20	37	54.4	653	1	LEPA_MYCTU	P17139 mycobacteri
21	37	54.4	659	1	NRTC_SYNP7	P38045 synechococc
22	36	52.9	229	1	UNG_CHLTR	O84613 chlamydia t
23	36	52.9	291	1	IBP3_BOVIN	P20959 bos taurus
24	36	52.9	352	1	TCBF_PSESQ	P27101 pseudomonas
25	36	52.9	352	1	TFTE_BURCE	Q45072 burkholderi
26	36	52.9	354	1	TFE1_ALCEU	P27137 alcaligenes
27	36	52.9	356	1	HXB2_HUMAN	P14652 homo sapien
28	36	52.9	1022	1	CA26_CHICK	P15988 gallus gall
29	36	52.9	1212	1	UTY_MOUSE	P79457 mus musculu
30	36	52.9	1337	1	YDM5_SCHPO	P87136 schizosacch
31	36	52.9	1572	1	BAI2_HUMAN	O60241 homo sapien
32	36	52.9	2124	1	Y192_HUMAN	Q93074 homo sapien
33	35.5	52.2	249	1	RU2A_ARATH	P43333 arabidopsis

34	35.5	52.2	501	1	GSPE_AERHY	P31741 aeromonas h
35	35	51.5	85	1	PLF4_SHEEP	P30035 ovis aries
36	35	51.5	136	1	SR19_ORYSA	P49964 oryza sativ
37	35	51.5	266	1	IBP3_PIG	P16611 sus scrofa
38	35	51.5	267	1	CYSH_PSEAE	O05927 pseudomonas
39	35	51.5	271	1	YPQQ_KLEPN	P27509 klebsiella
40	35	51.5	299	1	FCN3_HUMAN	O75636 homo sapien
41	35	51.5	319	1	RM04_YEAST	P36517 saccharomyc
42	35	51.5	330	1	ETFA_CLOTS	P71153 clostridium
43	35	51.5	348	1	YT35_STRFR	P20186 streptomyce
44	35	51.5	387	1	UL33_RCMVM	O12000 rat cytomeg
45	35	51.5	454	1	GLNA_HALVO	P43386 halobacteri
46	35	51.5	481	1	MSK_MOUSE	Q60670 mus musculu
47	35	51.5	543	1	PKAA_STRCO	P54739 streptomyce
48	35	51.5	553	1	DMWD_HUMAN	Q09019 homo sapien
49	35	51.5	573	1	YKK3_CAEEL	P34280 caenorhabdi
50	35	51.5	653	1	HGFA_MOUSE	Q9r098 mus musculu
51	35	51.5	704	1	RP3A_BOVIN	Q06846 bos taurus
52	35	51.5	714	1	HUNK_HUMAN	P57058 homo sapien
53	35	51.5	776	1	SNIL_RAT	Q9rlu5 rattus norv
54	35	51.5	778	1	TAST_HUMAN	Q12815 homo sapien
55	35	51.5	786	1	SNIL_HUMAN	P57059 homo sapien
56	35	51.5	1019	1	LFC_TACTR	P28175 tachyples
57	35	51.5	1029	1	CA26_MOUSE	Q02788 mus musculu
58	35	51.5	1162	1	LEPR_RAT	Q62959 rattus norv
59	35	51.5	2871	1	DESP_HUMAN	P15924 homo sapien
60	34	50.0	77	1	Y188_ARCFU	O30050 archaeoglob
61	34	50.0	121	1	SECR_HUMAN	P09683 homo sapien
62	34	50.0	162	1	PHAC_CYAPA	P16282 cyanophora
63	34	50.0	178	1	BTC_HUMAN	P35070 homo sapien
64	34	50.0	188	1	ADML_CANFA	O77559 canis famil
65	34	50.0	205	1	YK07_YEAST	P36061 saccharomyc

ALIGNMENTS

RESULT	1
NARQ_HAEIN	
ID	NARQ_HAEIN
AC	P44604; STANDARD; PRT; 567 AA.
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	SENSOR PROTEIN NARQ HOMOLOG (EC 2.7.3.-).
GN	NARQ OR HI0267.
OS	Haemophilus influenzae.
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC	Haemophilus.
OX	NCBI_TaxID=727;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=RD / KW20 / ATCC 51907;
RX	MEDLINE=95350630; PubMed=7542800;
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA	Venter J.C.;
RT	"Whole-genome random sequencing and assembly of Haemophilus
RT	influenzae Rd.";
RL	Science 269:496-512(1995).
CC	-!- FUNCTION: PROBABLE MEMBER OF A TWO-COMPONENT REGULATORY SYSTEM;
CC	HOWEVER IT IS NOT KNOWN WHAT PROTEIN IT PHOSPHORYLATES AND IN WHICH
CC	REGULATORY PATHWAY IT ACTS AS THE NARL AND OTHER NAR GENES DO NOT
CC	EXIST IN H. INFLUENZAE.
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC	(PROBABLE).
CC	-!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE

```
CC CC KINASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensee@isb-sib.ch).
CC -----
CC EMBL; U32713; AAC21933.1;
CC TIGR; H10267;
CC InterPro; IPR000410;
CC InterPro; IPR000658;
CC Pfam; PF00672; DUF5; 1.
CC Pfam; PF00512; signal; 1.
CC Sensory transduction; Transferrase; Kinase; Phosphorylation;
CC Transmembrane; Inner membrane.
CC DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 14 34 POTENTIAL.
CC DOMAIN 35 147 PERIPLASMIC (POTENTIAL).
CC TRANSMEM 148 172 POTENTIAL.
CC DOMAIN 173 567 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 350 567 TRANSFERASE DOMAIN (POTENTIAL).
CC MOD_RES 373 373 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC SEQUENCE 567 AA; 65238 MW; A74F54998C6ECAB9 CRC64;
CC
CC Query Match 67.6%; Score 46; DB 1; Length 567;
CC Best Local Similarity 66.7%; Pred. NO. 2.6;
CC Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 1 AGLLCPPDRPLE 12
CC Db 326 AGLPCPPDRIMO 337
CC
CC RESULT 2
CC YC64_MYCTU STANDARD; PRT; 397 AA.
CC ID Q11055;
CC AC Q11055; 1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 01-OCT-2000 (Rel. 40, Last annotation update)
CC DE HYPOTHETICAL 42.2 KDA PROTEIN RV1264.
CC GN RV1264 OR MTCY50.18C.
CC OS Mycobacterium tuberculosis.
CC OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
CC OX NCBI_TaxID=1773;
CC RN [1]
CC SEQUENCE FROM N.A.
CC STRAIN-H37RV;
CC MEDLINE-98295987; PubMed=9634230;
CC Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
CC Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
CC Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
CC Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
CC Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
CC Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
CC Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
CC Taylor K., Whitehead S., Barrell B.G.;
CC "Deciphering the biology of Mycobacterium tuberculosis from the
CC complete genome sequence."
CC Nature 393:537-544(1998).
CC -!- SIMILARITY: CONTAINS 1 GUANYLATE CYCLASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to licensee@isb-sib.ch).
CC -----
CC EMBL; 277137; CAB00890.1;
CC HSP; Q02846; 1AWL.
CC Tuberculin; RV1264;
CC Pfam; PF00211; guanylate_cyc; 2.
CC PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
CC KW Hypothetical protein.
CC SQ SEQUENCE 397 AA; 42232 MW; F6C212A181DB5AD3 CRC64;
CC
CC Query Match 64.7%; Score 44; DB 1; Length 397;
CC Best Local Similarity 77.8%; Pred. NO. 4;
CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 3 LLCPPDRPL 11
CC Db 269 LVCPDPAPL 277
CC
CC RESULT 3
CC V66K_BWYVF STANDARD; PRT; 607 AA.
CC ID P09506;
CC AC P09506;
CC DT 01-MAR-1989 (Rel. 10, Created)
CC DT 01-MAR-1989 (Rel. 10, Last sequence update)
CC DT 01-OCT-1989 (Rel. 12, Last annotation update)
CC DE 66.2 KDA PROTEIN (ORF 2).
CC OS Beet western yellow virus (isolate FL-1) (BWV).
CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
CC OC Polerovirus.
CC OX NCBI_TaxID=12043;
CC RN [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=89057523; PubMed=3194229;
CC Veidt I., Lot H., Leiser M., Scheidecker D., Guille H., Richards K.,
CC Jonard G.;
CC "Nucleotide sequence of beet western yellow virus RNA."
CC Nucleic Acids Res. 16:9917-9932(1988).
CC -!- SIMILARITY: TO POTATO LEAFROLL VIRUS ORF2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensee@isb-sib.ch).
CC -----
CC EMBL; X13063; CAA31463.1;
CC PIR; S01939; S01939.
CC InterPro; IPR000382;
CC Pfam; PF02122; Luteo_ORF2; 1.
CC PRINTS; PR00913; LVIRUSORF2.
CC SEQUENCE 607 AA; 66210 MW; 2A41F82911DEC84F CRC64;
CC
CC Query Match 60.3%; Score 41; DB 1; Length 607;
CC Best Local Similarity 85.7%; Pred. NO. 18;
CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 5 QPDRPL 11
CC Db 62 CPEPRPL 68
CC
CC RESULT 4
CC ERB2_HUMAN STANDARD; PRT; 1255 AA.
CC ID P04626;
CC AC P04626;
CC DT 13-AUG-1987 (Rel. 05, Created)
CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
CC DT 01-OCT-2000 (Rel. 40, Last annotation update)
```

DE RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2 PRECURSOR (EC 2.7.1.112)  
DE (P15ERBB2) (NEU PROTO-ONCOGENE) (C-ERBB-2) (TYROSINE KINASE-TYPE CELL  
DE SURFACE RECEPTOR HER2) (MLN 19).  
GN ERBB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139(1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=2995967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RT human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
CC ALPHA AND AMPHIREGULIN.  
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
CC PROTEIN TYROSINE PHOSPHATE.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC RESIDUES (BY SIMILARITY).  
CC -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN  
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M11767; AAA35808.1; -.  
DR EMBL; M11761; AAA35808.1; JOINED.  
DR EMBL; M11762; AAA35808.1; JOINED.  
DR EMBL; M11763; AAA35808.1; JOINED.  
DR EMBL; M11764; AAA35808.1; JOINED.  
DR EMBL; M11765; AAA35808.1; JOINED.  
DR EMBL; M11766; AAA35808.1; JOINED.  
DR EMBL; M11730; AAA75493.1; -.

DR EMBL; M12036; AAA35978.1; -.  
DR EMBL; X03363; CAA27060.1; -.  
DR PIR; A25491; A25491.  
DR PIR; A24571; A24571.  
DR HSSP; P11362; 1FGI.  
DR MIM; 164870; -.  
DR InterPro; IPR000494; -.  
DR InterPro; IPR000719; -.  
DR InterPro; IPR001245; -.  
DR InterPro; IPR002174; -.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF00069; pkinase; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Polymorphism.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 653 675 POTENTIAL.  
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 720 987 PROTEIN KINASE.  
FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
FT BINDING 753 753 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 654 654 I -> V.  
FT VARIANT 655 655 /FTid=VAR\_004077.  
FT VARIANT 655 655 I -> V.  
FT CONFLICT 1170 1170 /FTid=VAR\_004078.  
FT CONFLICT 1170 1170 P -> A (IN REF. 2).  
SQ SEQUENCE 1255 AA; 137909 MW; 39E9DFDA04DCF962 CRC64;  
  
Query Match 60.3%; Score 41; DB 1; Length 1255;  
Best Local Similarity 66.7%; Pred. No. 35;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 2 GLICPDPRP 10  
| | | | |  
Db 1029 GFFCPDPAP 1037  
  
RESULT 5  
DPAL\_ECOLI  
ID DPAL\_ECOLI STANDARD; PRT; 398 AA.  
AC Q46804;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PUTATIVE DIAMINOPROPIONATE AMMONIA-LYASE (EC 4.3.1.15)  
DE (DIAMINOPROPIONATASE) (ALPHA,BETA-DIAMINOPROPIONATE AMMONIA-LYASE).  
GN YGEX.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MGI655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,



RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474 (1997).  
 CC -!- FUNCTION: CATALYZES THE ALPHA,BETA-ELIMINATION REACTION OF BOTH L-  
 CC AND D-ALPHA,BETA-DIAMINOPROPIONATE, THE MOST SUITABLE SUBSTRATES  
 CC TO FORM PYRUVATE AND AMMONIA. THE L- AND D-ISOMERS OF SERINE ARE  
 CC ALSO DEGRADED, THOUGH SLOWLY; IT IS THE ONLY SERINE DEHYDRATASE  
 CC WHICH CAN ELIMINATE AN AMINO GROUP AT THE BETA-CARBON POSITION (BY  
 CC SIMILARITY).  
 CC CATALYTIC ACTIVITY: 2,3-DIAMINOPROPIONATE = PYRUVATE + 2 NH(3).  
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -!- SIMILARITY: TO S-TYPHIMURIUM DIAMINOPROPIONATE AMMONIA-LYASE AND  
 CC TO THREONINE DEHYDRATASE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U28375; AAA83052.1; -  
 CC EMBL; AE000370; AAC75909.1; -  
 CC EcoGene; EGI3054; Ygex.  
 CC InterPro; IPR001926; -  
 CC Pfam; PF00291; S-T-dehydratase; 1.  
 CC Hypothetical protein; 77; PYRIDOXAL PHOSPHATE (POTENTIAL).  
 CC BINDING 77  
 CC SEQUENCE 398 AA; 43327 MW; FAF1277E86D60232 CRC64;  
 CC SQ  
 Query Match 58.8%; Score 40; DB 1; Length 398;  
 Best Local Similarity 63.6%; Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 AGLLCPDPRPL 11  
 DB 287 AGLACGEPNPL 297  
 RESULT 6  
 ID UAPA\_EME1 STANDARD; PRT; 615 AA.  
 AC Q07307; 1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE URIC ACID-XANTHINE PERMEASE (UAPA TRANSPORTER).  
 GN UAPA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=5072;  
 RN [1]  
 RP SEQUENCE FROM N.A. PubMed=8226862;  
 RX MEDLINE=94043131; Diallinas G., Scazzocchio C.,  
 RA Gorfinkel L., Diallinas G., Scazzocchio C.,  
 RT "Sequence and regulation of the uapa gene encoding a uric acid-  
 RL xanthine permease in the fungus Aspergillus nidulans.";  
 RL J. Biol. Chem. 268:23376-23381(1993).  
 RN [2]  
 RP REVISIONS.  
 RA Diallinas G.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- INDUCTION: INDUCIBLE BY 2-THIOURIC ACID, AND HIGHLY REPRESSIBLE BY  
 CC AMMONIUM.  
 CC -!- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X71807; CAA50681.1; -  
 CC PIR; S36031; S36031.  
 CC InterPro; IPR000444; -  
 CC Pfam; PF00860; xanthur\_permease; 1.  
 CC PROSITE; PS01116; XANTH\_URACIL\_PERMEASE; 1.  
 CC TRANSPORT; Transmembrane; Glycoprotein.  
 CC TRANSMEM 118 138  
 CC TRANSMEM 152 172  
 CC TRANSMEM 182 202  
 CC TRANSMEM 229 250  
 CC TRANSMEM 258 278  
 CC TRANSMEM 305 325  
 CC TRANSMEM 337 356  
 CC TRANSMEM 379 402  
 CC TRANSMEM 468 488  
 CC TRANSMEM 492 512  
 CC TRANSMEM 523 543  
 CC TRANSMEM 563 583  
 CC CARBOHYD 10  
 CC SEQUENCE 615 AA; 65453 MW; E56A984D956897E7 CRC64;  
 CC SQ  
 Query Match 58.1%; Score 39.5; DB 1; Length 615;  
 Best Local Similarity 61.5%; Pred. No. 32;  
 Matches 8; Conservative 1; Mismatches 1; Indels 3; Gaps 1;  
 QY 2 GLLCPD---PRPL 11  
 DB 293 GMLCPSATAPRPL 305  
 RESULT 7  
 ID IMA\_LYCES STANDARD; PRT; 527 AA.  
 AC O22478; 1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 37, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE IMPORTIN ALPHA SUBUNIT (KARYOPHERIN ALPHA SUBUNIT) (KAP ALPHA).  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kunik T., Mizrachi L., Citovsky V., Gafni Y.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: BINDS SPECIFICALLY AND DIRECTLY TO SUBSTRATES CONTAINING  
 CC EITHER A SIMPLE OR BIPARTITE NLS MOTIF. PROMOTES DOCKING OF IMPORT  
 CC SUBSTRATES TO THE NUCLEAR ENVELOPE. SEEMS TO ACT AS A CYTOSOLIC  
 CC RECEPTOR FOR BOTH SIMPLE AND BIPARTITE NLS MOTIFS (BY SIMILARITY).  
 CC -!- SUBUNIT: FORMS A COMPLEX WITH IMPORTIN BETA-1 SUBUNIT (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE IMPORTIN ALPHA FAMILY.  
 CC -!- SIMILARITY: CONTAINS 8 ARM REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

```
DR EMBL; AF017252; AAC23722.1; -.
DR InterPro; IPR000225; -.
DR InterPro; IPR002652; -.
DR Pfam; PF00514; Armadillo_seg; 8.
DR Pfam; PF01749; IBB; 1.
DR PROSITE; PS50176; ARM_REPEAT; 5.
KW Transport; Protein transport; Repeat.
FT DOMAIN 12 51 IBB.
FT REPEAT 109 151 ARM 1.
FT REPEAT 152 196 ARM 2.
FT REPEAT 197 234 ARM 3.
FT REPEAT 235 279 ARM 4.
FT REPEAT 280 319 ARM 5.
FT REPEAT 320 362 ARM 6.
FT REPEAT 363 403 ARM 7.
FT REPEAT 403 445 ARM 8.
FT DOMAIN 446 527 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 527 AA; 58605 MW; 4A3F01691CEF4817 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 527;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
Db 422 LVCPDPR 428

RESULT 8
IMAL_ARATH
ID IMAL_ARATH STANDARD; PRT; 596 AA.
AC Q96321;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE IMPORTIN ALPHA-1 SUBUNIT (KARYOPHERIN ALPHA-1 SUBUNIT) (KAP ALPHA).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Ballas N., Citovsky V.;
RT "AtKAPalpha gene from Arabidopsis encodes a protein that mediates
RT nuclear import of Agrobacterium VirD2 protein.";
RL (In) Plant Gene Register PGR97-129.
CC -!- FUNCTION: BINDS SPECIFICALLY AND DIRECTLY TO SUBSTRATES CONTAINING
CC EITHER A SIMPLE OR BIPARTITE NLS MOTIF. PROMOTES DOCKING OF IMPORT
CC SUBSTRATES TO THE NUCLEAR ENVELOPE. SEEMS TO ACT AS A CYTOSOLIC
CC RECEPTOR FOR BOTH SIMPLE AND BIPARTITE NLS MOTIFS (BY SIMILARITY).
CC CELLULAR RECEPTOR FOR THE NUCLEAR IMPORT OF THE VIRD2 PROTEIN OF
CC AGROBACTERIUM.
CC -!- SUBUNIT: FORMS A COMPLEX WITH IMPORTIN BETA-1 SUBUNIT (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE IMPORTIN ALPHA FAMILY.
CC -!- SIMILARITY: CONTAINS 8 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U69533; AAB72116.1; -.
CC HSSP; Q02248; 2BCT.
DR InterPro; IPR000225; -.
DR InterPro; IPR002652; -.
DR Pfam; PF00514; Armadillo_seg; 8.
```

```
DR Pfam; PF01749; IBB; 1.
DR PROSITE; PS50176; ARM_REPEAT; 3.
KW Transport; Protein transport; Repeat.
FT DOMAIN 12 51 IBB.
FT REPEAT 109 151 ARM 1.
FT REPEAT 152 196 ARM 2.
FT REPEAT 197 234 ARM 3.
FT REPEAT 235 279 ARM 4.
FT REPEAT 280 319 ARM 5.
FT REPEAT 320 362 ARM 6.
FT REPEAT 363 403 ARM 7.
FT REPEAT 403 445 ARM 8.
FT DOMAIN 446 596 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 596 AA; 65606 MW; 2A2689E1C28F43E7 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 596;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
Db 422 LVCPDPR 428

RESULT 9
EGF_HUMAN
ID EGF_HUMAN STANDARD; PRT; 1207 AA.
AC P01133;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PRO-EPIDERMAL GROWTH FACTOR PRECURSOR (EGF) [CONTAINS: EPIDERMAL
DE GROWTH FACTOR (UROGASTRONE)].
GN EGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=87066721; PubMed=3491360;
RA Bell G.I., Fong N.M., Stempien M.M., Wormsted M.A., Caput D.,
RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;
RT "Human epidermal growth factor precursor: cDNA sequence, expression
RT in vitro and gene organization.";
RL Nucleic Acids Res. 14:8427-8446(1986).
RN [2]
RP SEQUENCE OF 971-1023.
RX MEDLINE=77117897; PubMed=300079;
RA Gregory H., Preston B.M.;
RT "The primary structure of human urogastrone.";
RL Int. J. Pept. Protein Res. 9:107-118(1977).
RN [3]
RP SEQUENCE OF 971-1023.
RX MEDLINE=89391964; PubMed=2789514;
RA Furuya M., Akashi S., Hirayama K.;
RT "The primary structure of human EGF produced by genetic engineering,
RT studied by high-performance tandem mass spectrometry.";
RL Biochem. Biophys. Res. Commun. 163:1100-1106(1989).
RN [4]
RP STRUCTURE BY NMR OF EGF.
RX MEDLINE=92395667; PubMed=1522591;
RA Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;
RT "Human epidermal growth factor. High resolution solution structure
RT and comparison with human transforming growth factor alpha.";
RL J. Mol. Biol. 227:271-282(1992).
CC -!- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X04571; CAA28240.1; -  
DR PIR; A25531; EGHU.  
DR HSSP; P01132; IEPH.  
DR MIM; 131530; -  
DR InterPro; IPR000033; -  
DR InterPro; IPR000152; -  
DR InterPro; IPR000561; -  
DR InterPro; IPR001336; -  
DR InterPro; IPR001881; -  
DR Pfam; PF00008; EGF; 9.  
DR Pfam; PF00058; ldl\_recept\_b; 7.  
DR PRINTS; PR00009; EGF\_TGF.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 7.  
DR PROSITE; PS01187; EGF\_CA; 3.  
KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;  
KW Signal; Polymorphism.  
FT SIGNAL 1 22  
FT CHAIN 23 1207  
FT CHAIN 23 1023  
FT DOMAIN 23 1032  
FT TRANSMEM 1033 1053  
FT DOMAIN 1054 1207  
FT DOMAIN 314 355  
FT DOMAIN 356 396  
FT DOMAIN 397 437  
FT DOMAIN 438 477  
FT DOMAIN 478 518  
FT DOMAIN 519 559  
FT DOMAIN 560 600  
FT DOMAIN 601 641  
FT DOMAIN 642 682  
FT DOMAIN 683 723  
FT DOMAIN 724 764  
FT DOMAIN 765 805  
FT DOMAIN 806 846  
FT DOMAIN 847 887  
FT DOMAIN 888 928  
FT DOMAIN 929 969  
FT DOMAIN 970 1010  
FT DOMAIN 1011 1051  
FT DOMAIN 1052 1092  
FT DOMAIN 1093 1133  
FT DOMAIN 1134 1174  
FT DOMAIN 1175 1215  
FT DOMAIN 1216 1256  
FT DOMAIN 1257 1297  
FT DOMAIN 1298 1338  
FT DOMAIN 1339 1379  
FT DOMAIN 1380 1420  
FT DOMAIN 1421 1461  
FT DOMAIN 1462 1502  
FT DOMAIN 1503 1543  
FT DOMAIN 1544 1584  
FT DOMAIN 1585 1625  
FT DOMAIN 1626 1666  
FT DOMAIN 1667 1707  
FT DOMAIN 1708 1748  
FT DOMAIN 1749 1789  
FT DOMAIN 1790 1830  
FT DOMAIN 1831 1871  
FT DOMAIN 1872 1912  
FT DOMAIN 1913 1953  
FT DOMAIN 1954 1994  
FT DOMAIN 1995 2035  
FT DOMAIN 2036 2076  
FT DOMAIN 2077 2117  
FT DOMAIN 2118 2158  
FT DOMAIN 2159 2199  
FT DOMAIN 2200 2240  
FT DOMAIN 2241 2281  
FT DOMAIN 2282 2322  
FT DOMAIN 2323 2363  
FT DOMAIN 2364 2404  
FT DOMAIN 2405 2445  
FT DOMAIN 2446 2486  
FT DOMAIN 2487 2527  
FT DOMAIN 2528 2568  
FT DOMAIN 2569 2609  
FT DOMAIN 2610 2650  
FT DOMAIN 2651 2691  
FT DOMAIN 2692 2732  
FT DOMAIN 2733 2773  
FT DOMAIN 2774 2814  
FT DOMAIN 2815 2855  
FT DOMAIN 2856 2896  
FT DOMAIN 2897 2937  
FT DOMAIN 2938 2978  
FT DOMAIN 2979 3019  
FT DOMAIN 3020 3060  
FT DOMAIN 3061 3101  
FT DOMAIN 3102 3142  
FT DOMAIN 3143 3183  
FT DOMAIN 3184 3224  
FT DOMAIN 3225 3265  
FT DOMAIN 3266 3306  
FT DOMAIN 3307 3347  
FT DOMAIN 3348 3388  
FT DOMAIN 3389 3429  
FT DOMAIN 3430 3470  
FT DOMAIN 3471 3511  
FT DOMAIN 3512 3552  
FT DOMAIN 3553 3593  
FT DOMAIN 3594 3634  
FT DOMAIN 3635 3675  
FT DOMAIN 3676 3716  
FT DOMAIN 3717 3757  
FT DOMAIN 3758 3798  
FT DOMAIN 3799 3839  
FT DOMAIN 3840 3880  
FT DOMAIN 3881 3920  
FT DOMAIN 3921 3961  
FT DOMAIN 3962 4002  
FT DOMAIN 4003 4043  
FT DOMAIN 4044 4084  
FT DOMAIN 4085 4125  
FT DOMAIN 4126 4166  
FT DOMAIN 4167 4207  
FT DOMAIN 4208 4248  
FT DOMAIN 4249 4289  
FT DOMAIN 4290 4330  
FT DOMAIN 4331 4371  
FT DOMAIN 4372 4412  
FT DOMAIN 4413 4453  
FT DOMAIN 4454 4494  
FT DOMAIN 4495 4535  
FT DOMAIN 4536 4576  
FT DOMAIN 4577 4617  
FT DOMAIN 4618 4658  
FT DOMAIN 4659 4699  
FT DOMAIN 4700 4740  
FT DOMAIN 4741 4781  
FT DOMAIN 4782 4822  
FT DOMAIN 4823 4863  
FT DOMAIN 4864 4904  
FT DOMAIN 4905 4945  
FT DOMAIN 4946 4986  
FT DOMAIN 4987 5027  
FT DOMAIN 5028 5068  
FT DOMAIN 5069 5109  
FT DOMAIN 5110 5150  
FT DOMAIN 5151 5191  
FT DOMAIN 5192 5232  
FT DOMAIN 5233 5273  
FT DOMAIN 5274 5314  
FT DOMAIN 5315 5355  
FT DOMAIN 5356 5396  
FT DOMAIN 5397 5436  
FT DOMAIN 5437 5477  
FT DOMAIN 5478 5518  
FT DOMAIN 5519 5559  
FT DOMAIN 5560 5600  
FT DOMAIN 5601 5641  
FT DOMAIN 5642 5682  
FT DOMAIN 5683 5723  
FT DOMAIN 5724 5764  
FT DOMAIN 5765 5805  
FT DOMAIN 5806 5846  
FT DOMAIN 5847 5887  
FT DOMAIN 5888 5928  
FT DOMAIN 5929 5969  
FT DOMAIN 5970 6010  
FT DOMAIN 6011 6051  
FT DOMAIN 6052 6092  
FT DOMAIN 6093 6133  
FT DOMAIN 6134 6174  
FT DOMAIN 6175 6215  
FT DOMAIN 6216 6256  
FT DOMAIN 6257 6297  
FT DOMAIN 6298 6338  
FT DOMAIN 6339 6378  
FT DOMAIN 6379 6419  
FT DOMAIN 6420 6460  
FT DOMAIN 6461 6501  
FT DOMAIN 6502 6542  
FT DOMAIN 6543 6583  
FT DOMAIN 6584 6624  
FT DOMAIN 6625 6665  
FT DOMAIN 6666 6706  
FT DOMAIN 6707 6747  
FT DOMAIN 6748 6788  
FT DOMAIN 6789 6829  
FT DOMAIN 6830 6870  
FT DOMAIN 6871 6910  
FT DOMAIN 6911 6951  
FT DOMAIN 6952 6992  
FT DOMAIN 6993 7033  
FT DOMAIN 7034 7074  
FT DOMAIN 7075 7115  
FT DOMAIN 7116 7156  
FT DOMAIN 7157 7196  
FT DOMAIN 7197 7237  
FT DOMAIN 7238 7278  
FT DOMAIN 7279 7319  
FT DOMAIN 7320 7360  
FT DOMAIN 7361 7401  
FT DOMAIN 7402 7442  
FT DOMAIN 7443 7483  
FT DOMAIN 7484 7524  
FT DOMAIN 7525 7565  
FT DOMAIN 7566 7606  
FT DOMAIN 7607 7647  
FT DOMAIN 7648 7688  
FT DOMAIN 7689 7729  
FT DOMAIN 7730 7770  
FT DOMAIN 7771 7811  
FT DOMAIN 7812 7852  
FT DOMAIN 7853 7893  
FT DOMAIN 7894 7934  
FT DOMAIN 7935 7975  
FT DOMAIN 7976 8016  
FT DOMAIN 8017 8057  
FT DOMAIN 8058 8098  
FT DOMAIN 8099 8139  
FT DOMAIN 8140 8180  
FT DOMAIN 8181 8221  
FT DOMAIN 8222 8262  
FT DOMAIN 8263 8303  
FT DOMAIN 8304 8344  
FT DOMAIN 8345 8385  
FT DOMAIN 8386 8426  
FT DOMAIN 8427 8467  
FT DOMAIN 8468 8508  
FT DOMAIN 8509 8549  
FT DOMAIN 8550 8590  
FT DOMAIN 8591 8631  
FT DOMAIN 8632 8672  
FT DOMAIN 8673 8713  
FT DOMAIN 8714 8754  
FT DOMAIN 8755 8795  
FT DOMAIN 8796 8836  
FT DOMAIN 8837 8877  
FT DOMAIN 8878 8918  
FT DOMAIN 8919 8959  
FT DOMAIN 8960 9000  
FT DOMAIN 9001 9041  
FT DOMAIN 9042 9082  
FT DOMAIN 9083 9123  
FT DOMAIN 9124 9164  
FT DOMAIN 9165 9205  
FT DOMAIN 9206 9246  
FT DOMAIN 9247 9287  
FT DOMAIN 9288 9328  
FT DOMAIN 9329 9369  
FT DOMAIN 9370 9410  
FT DOMAIN 9411 9451  
FT DOMAIN 9452 9492  
FT DOMAIN 9493 9533  
FT DOMAIN 9534 9574  
FT DOMAIN 9575 9615  
FT DOMAIN 9616 9656  
FT DOMAIN 9657 9697  
FT DOMAIN 9698 9738  
FT DOMAIN 9739 9779  
FT DOMAIN 9780 9820  
FT DOMAIN 9821 9861  
FT DOMAIN 9862 9902  
FT DOMAIN 9903 9943  
FT DOMAIN 9944 9984  
FT DOMAIN 9985 10000  
FT CARBOHYD 38  
FT CARBOHYD 104  
FT CARBOHYD 117  
FT CARBOHYD 148  
FT CARBOHYD 148

FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 708 708 I -> M.  
FT SEQUENCE 1207 AA; 133945 MW; D627DC828EF782E9 CRC64;  
SQ  
Query Match 57.4%; Score 39; DB 1; Length 1207;  
Best Local Similarity 66.7%; Pred. No. 72;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GLICPDPRP 10  
DB 948 GLICPDSTP 956  
RESULT 10  
Z142 HUMAN  
ID Z142 HUMAN STANDARD; PRT; 1687 AA.  
AC P52746; Q92510;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ZINC FINGER PROTEIN 142 (HA4654).  
GN ZNF142 OR KIAA0236.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=97191544; PubMed=9039502;  
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y.,  
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.,  
RT "prediction of the coding sequences of unidentified human genes. VI.  
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
RT analysis of cDNA clones from cell line KG-1 and brain.";  
RL DNA Res. 3:321-329(1996).  
RN [2]  
RP SEQUENCE OF 460-632 FROM N.A.  
RC TISSUE=Insulinoma;  
RX MEDLINE=96044430; PubMed=7557990;  
RA Tommerup N., Vissing H.;  
RT "Isolation and fine mapping of 16 novel human zinc finger-encoding  
RT cDNAs identify putative candidate genes for developmental and  
RT malignant disorders.";  
RL Genomics 27:259-264(1995).  
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D87073; BAA13242.1; -  
CC EMBL; U09849; AAC50265.1; -  
CC HSSP; P25490; IUBD.  
CC MIM; 604083; -  
CC InterPro; IPR000822; -  
CC Pfam; PF00096; zf-C2H2; 35.  
CC PRINTS; PR00048; ZINCFINGER.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 17.  
CC PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 18.  
DR

KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
KW Nuclear protein; Repeat.  
FT DOMAIN 163 1643 ZINC FINGERS.  
FT ZN\_FING 163 185 C2H2-TYPE.  
FT ZN\_FING 219 242 C2H2-TYPE.  
FT ZN\_FING 253 275 C2H2-TYPE.  
FT ZN\_FING 286 311 C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 316 340 C2H2-TYPE.  
FT ZN\_FING 343 366 C2H2-TYPE.  
FT ZN\_FING 372 395 C2H2-TYPE.  
FT ZN\_FING 401 423 C2H2-TYPE.  
FT ZN\_FING 429 451 C2H2-TYPE.  
FT ZN\_FING 457 479 C2H2-TYPE.  
FT ZN\_FING 485 507 C2H2-TYPE.  
FT ZN\_FING 512 536 C2H2-TYPE.  
FT ZN\_FING 544 567 C2H2-TYPE.  
FT ZN\_FING 573 596 C2H2-TYPE.  
FT ZN\_FING 1135 1158 C2H2-TYPE.  
FT ZN\_FING 1171 1194 C2H2-TYPE.  
FT ZN\_FING 1200 1222 C2H2-TYPE.  
FT ZN\_FING 1228 1251 C2H2-TYPE.  
FT ZN\_FING 1257 1280 C2H2-TYPE.  
FT ZN\_FING 1286 1309 C2H2-TYPE.  
FT ZN\_FING 1328 1351 C2H2-TYPE.  
FT ZN\_FING 1354 1377 C2H2-TYPE.  
FT ZN\_FING 1380 1403 C2H2-TYPE.  
FT ZN\_FING 1424 1446 C2H2-TYPE.  
FT ZN\_FING 1452 1474 C2H2-TYPE.  
FT ZN\_FING 1480 1502 C2H2-TYPE.  
FT ZN\_FING 1508 1530 C2H2-TYPE.  
FT ZN\_FING 1536 1559 C2H2-TYPE.  
FT ZN\_FING 1565 1587 C2H2-TYPE.  
FT ZN\_FING 1593 1615 C2H2-TYPE.  
FT ZN\_FING 1621 1643 C2H2-TYPE.  
FT CONFLICT 460 463  
FT CONFLICT 541 541  
FT CONFLICT 541 541  
FT SEQUENCE 1687 AA; 187865 MW; 42847520F04D38E7 CRC64;  
  
Query Match 57.4%; Score 39; DB 1; Length 1687;  
Best Local Similarity 70.0%; Pred. No. 97;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY - 2 GLLCPDRPL 11  
Db 1334 GLLCPSPASL 1343  
  
RESULT 11  
ETFA\_MEGEL STANDARD; PRT; 338 AA.  
AC O85692;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNIT (ALPHA-ETF) (ELECTRON  
DE TRANSFER FLAVOPROTEIN LARGE SUBUNIT) (ETFLS).  
GN ETFA.  
OS Megaspheara elsdenii.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Sporomusa subbranch;  
OC Megaspheara.  
OX NCBI\_TaxID=907;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LC1;  
RX MEDLINE=98361974; PubMed=9694853;  
RA O'Neill H., Mayhew S.G., Butler G.;  
RT "Cloning and analysis of the genes for a novel electron-transferring  
RT flavoprotein from Megaspheara elsdenii. Expression and  
RT characterization of the recombinant protein.\*;  
RL J. Biol. Chem. 273:21015-21024(1998).  
CC -!- FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC  
CC ELECTRON ACCEPTOR FOR OTHER DEHYDROGENASES. IT TRANSFERS THE

CC ELECTRONS TO THE MAIN RESPIRATORY CHAIN VIA ETF-UBIQUINONE  
CC OXIDOREDUCTASE (ETF DEHYDROGENASE) (BY SIMILARITY).  
CC -!- COFACTOR: CONTAINS TWO MOLECULES OF FAD PER DIMER.  
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.  
CC -!- SIMILARITY: BELONGS TO THE ETF ALPHA-SUBUNIT / FIXB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF072475; AAC31170.1; -.  
CC InterPro; IPR001308; -.  
CC Pfam; PF00766; ETF\_alpha; 1.  
DR PROSITE; PS00696; ETF\_ALPHA; FALSE\_NEG.  
KW Electron transport; Flavoprotein; FAD.  
FT NP\_BIND 275 303 FAD (ADP PART) (POTENTIAL).  
SQ SEQUENCE 338 AA; 36124 MW; 63FBD4CCF111AE77 CRC64;  
  
Query Match 55.9%; Score 38; DB 1; Length 338;  
Best Local Similarity 60.0%; Pred. No. 32;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AGLLCPDPRP 10  
Db 150 AEIICPDNRP 159  
  
RESULT 12  
AAT\_SYNY3  
ID AAT\_SYNY3 STANDARD; PRT; 389 AA.  
AC Q55128;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1) (TRANSAMINASE A) (ASPAT).  
GN ASPC OR SLL0402.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugiura M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
CC -!- CATALYTIC ACTIVITY: L-ASPARTATE + 2-OXOGLUTARATE = OXALOACETATE +  
CC L-GLUTAMATE.  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D64001; BAA10261.1; -.  
CC InterPro; IPR001176; -.  
DR PROSITE; PS00155; aminotran\_1; 1.  
DR PRINTS; PRO0753; ACCSYNTHASE.  
DR PROSITE; PS00105; AA\_TRANSFER\_CLASS\_1; 1.  
DR





RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
RA Zaccaria P., Devan M., Wilson R.K., de la Bastide M., Habermann K.,  
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
RA Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
RA Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
RA Chen E., Marra M., Martienssen R., McCombie W.R.;  
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
thaliana.";  
RL Nature 402:769-777(1999).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=94368830; PubMed=8086440;  
RA Horn M.A., Walker J.C.;  
RT "Biochemical properties of the autophosphorylation of RLK5, a  
receptor-like protein kinase from Arabidopsis thaliana.";  
RL Biochim. Biophys. Acta 1208:65-74(1994).  
CC -!- COFACTOR: HAVE SIGNIFICANTLY GREATER ACTIVITY IN THE PRESENCE OF  
MN2+ THAN MG2+.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ROOTS AND ROSETTES.  
CC -!- PTM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.  
CC -!- SIMILARITY: TO PROTEIN KINASES IN THE C-TERMINUS. BUT DOES NOT  
SEEM TO HAVE CONSERVED A KINASE ACTIVITY.  
CC -!- SIMILARITY: CONTAINS 18 LEUCINE-RICH REPEATS (LRR).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M84660; AAA32859.1; -.  
DR EMBL; AL021749; CAA16889.1; -.  
DR EMBL; AL161572; CAB79651.1; -.  
DR HSSP; P00523; 2PTK.  
DR InterPro; IPR000719; -.  
DR InterPro; IPR001611; -.  
DR InterPro; IPR002290; -.  
DR Pfam; PF00560; LRR; 14.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PR00019; LEURICHRPT.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Transmembrane; Glycoprotein; Phosphorylation; Leucine-rich repeat;  
KW Repeat; Signal.  
FT SIGNAL 1 14 POTENTIAL.  
FT CHAIN 15 999 RECEPTOR-LIKE PROTEIN KINASE 5.  
FT DOMAIN 15 621 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 622 641 POTENTIAL.  
FT DOMAIN 642 999 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 88 112 LRR 1.  
FT REPEAT 114 138 LRR 2.  
FT REPEAT 138 161 LRR 3.  
FT REPEAT 163 186 LRR 4.  
FT REPEAT 188 211 LRR 5.  
FT REPEAT 235 261 LRR 6.  
FT REPEAT 263 283 LRR 7.  
FT REPEAT 283 306 LRR 8.  
FT REPEAT 306 330 LRR 9.  
FT REPEAT 332 353 LRR 10.  
FT REPEAT 354 378 LRR 11.

FT REPEAT 402 426 LRR 12.  
FT REPEAT 426 450 LRR 13.  
FT REPEAT 452 474 LRR 14.  
FT REPEAT 498 522 LRR 15.  
FT REPEAT 524 547 LRR 16.  
FT REPEAT 549 567 LRR 17.  
FT REPEAT 567 592 LRR 18.  
FT DOMAIN 583 968 PROTEIN KINASE.  
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT NP\_BIND 689 697 ATP (BY SIMILARITY).  
FT BINDING 711 711 ATP (BY SIMILARITY).  
FT ACT\_SITE 819 819 BY SIMILARITY.  
FT MUTAGEN 711 711 K->E: LOSS OF CATALYTIC ACTIVITY.  
SQ SEQUENCE 999 AA; 109095 MW; F5793D899EA0C6A7 CRC64;  
  
Query Match 55.9%; Score 38; DB 1; Length 999;  
Best Local Similarity 70.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 GLLCPDPRPL 11  
IIII I I I  
Db 943 GLLCTSPPL 952  
  
RESULT 15  
YGN3\_YEAST  
ID YGN3\_YEAST STANDARD; PRT; 1264 AA.  
AC P53125;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE HYPOTHETICAL 145.6 KDA PROTEIN IN RPL1B-CEG1 INTERGENIC REGION.  
GN YGL133W OR G2842.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1679;  
RX MEDLINE=96437978; PubMed=8840506;  
RA Escribano V., Eraso P., Portillo F., Mazon M.J.;  
RT "Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces  
cerevisiae chromosome VII reveals SEC27, SSM1b, a putative  
S-adenosylmethionine-dependent enzyme and six new open reading  
frames.";  
RT Yeast 12:887-892(1996).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; 272655; CAA96844.1; -.  
DR SGD; S0003101; ITC1.  
KW Hypothetical protein.  
SQ SEQUENCE 1264 AA; 145642 MW; 45E4CF8835C7C746 CRC64;  
  
Query Match 55.9%; Score 38; DB 1; Length 1264;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11  
Db 9 ILLPDPKPL 17

RESULT 16  
P53\_CAVPO STANDARD; PRT; 391 AA.  
ID P53\_CAVPO  
AC Q9WUR6;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CELLULAR TUMOR ANTIGEN P53 (TUMOR SUPPRESSOR P53).  
GN TP53.  
OS Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
OX NCBI\_TaxID=10141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Spleen;  
RC MEDLINE=99265972; PubMed=10331945;  
RA D'Erchia A.M., Pesole G., Tullio A., Saccone C., Sbisa E.;  
RT "Guinea pig p53 mRNA: identification of new elements in coding and  
RT untranslated regions and their functional and evolutionary  
RL implications.";  
RL Genomics 58:50-64(1999).  
CC 1- FUNCTION: SUPPRESSOR IN MANY TUMOR TYPES. INDUCES  
CC GROWTH ARREST OR APOPTOSIS DEPENDING ON THE PHYSIOLOGICAL  
CC CIRCUMSTANCES OR CELL TYPE, BUT BOTH ACTIVITIES ARE INVOLVED IN  
CC TUMOR SUPPRESSION. ACTS IN CELL CYCLE REGULATION, IS A TRANS-  
CC ACTIVATOR THAT ACTS TO NEGATIVELY REGULATE CELL DIVISION BY  
CC CONTROLLING A SET OF GENES REQUIRED FOR THIS PROCESS. ONE OF THE  
CC ACTIVATED GENES IS AN INHIBITOR OF CYCLIN-DEPENDENT KINASES.  
CC APOPTOSIS INDUCTION SEEMS TO BE MEDIATED EITHER BY STIMULATION OF  
CC BAX AND FAS ANTIGEN EXPRESSION, OR BY REPRESSION OF BCL-2  
CC EXPRESSION.  
CC 1- SUBUNIT: BINDS DNA AS AN HOMOTETRAMER (BY SIMILARITY).  
CC 1- SUBCELLULAR LOCATION: NUCLEAR.  
CC 1- DISEASE: P53 IS FOUND IN INCREASED AMOUNTS IN A WIDE VARIETY  
CC OF TRANSFORMED CELLS. P53 IS FREQUENTLY MUTATED OR INACTIVATED  
CC IN MANY TYPES OF CANCER.  
CC 1- SIMILARITY: BELONGS TO THE P53 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ009673; CAB43196.1; --  
CC InterPro; IPR002117; --  
CC Pfam; PF00870; P53; 1.  
CC PRINTS; PR00386; P53SUPPRESSR.  
CC PROSITE; PS00348; P53; 1.  
CC Anti-oncogene; DNA-binding; Transcription regulation; Activator;  
KW Nuclear protein; DNA-binding; Phosphorylation; Apoptosis.  
FT DOMAIN 1 44 TRANSCRIPTION ACTIVATION (ACIDIC).  
FT DNA\_BIND 100 290 BY SIMILARITY.  
FT DOMAIN 323 354 OLIGOMERIZATION.  
FT DOMAIN 366 385 BASIC (REPRESSION OF DNA-BINDING).  
FT DOMAIN 309 321 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT MOD\_RES 390 390 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 391 AA; 43287 MW; 321D40702383573E CRC64;

Query Match 54.4%; Score 37; DB 1; Length 391;  
Best Local Similarity 66.7%; Pred. No. 53;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLCPDPRP 10  
Db 291 GGLCPPEPTP 299

RESULT 17  
REP1\_KLULA STANDARD; PRT; 415 AA.  
ID REP1\_KLULA  
AC PI3775;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE TRANS-ACTING FACTOR B (REP1).  
GN B.  
OS Kluyveromyces lactis (Yeast).  
OG Plasmid pKDI.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
OX NCBI\_TaxID=28985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86232585; PubMed=3520486;  
RA Chen X.J., Sallio M., Falcone C., Bianchi M.M., Fukuhara H.;  
RT "Sequence organization of the circular plasmid pKDI from the yeast  
RT Kluyveromyces drosophilarius.";  
RL Nucleic Acids Res. 14:4471-4481(1986).  
CC 1- FUNCTION: PLASMID PARTITION REQUIRE REP1, REP2, AND A CIS-ACTING  
CC DNA SEQUENCE (KNOWN AS STB). REP 1 MAY ACT BY INTERCALATING IN  
CC THE YEAST NUCLEAR MATRIX AND BINDING STB EITHER DIRECTLY OR VIA  
CC REP2.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X03961; CAA27593.1; --  
CC PIR; S28088; S28088.  
CC Plasmid; Trans-acting factor.  
KW SEQUENCE 415 AA; 48000 MW; D2800957245AE897 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 415;  
Best Local Similarity 70.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 LLCPPDRPLE 12  
Db 132 LLCPSPELLE 141

RESULT 18  
BAR2\_SCHCO STANDARD; PRT; 518 AA.  
ID BAR2\_SCHCO  
AC Q05659;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PHEROMONE B ALPHA 2 RECEPTOR (FRAGMENT).  
GN BAR2.  
OS Schizophyllum commune (Bracket fungus).  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Stereales;  
OC Schizophyllaceae; Schizophyllum.  
OX NCBI\_TaxID=5334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=12-43;  
RX MEDLINE=96080162; PubMed=7489716;  
RA Wendland J., Vaillancourt L.J., Hegner J., Lengeler K.B.,



RA Laddison K.J., Specht C.A., Raper C.A., Kothe E.;  
RT "The mating-type locus B alpha 1 of Schizophyllum commune contains a  
RL pheromone receptor gene and putative pheromone genes.";  
RL EMBO J. 14:5271-5278(1995).  
CC -!- FUNCTION: RECEPTOR FOR THE BAP2 PHEROMONE, A PRENYLATED MATING  
CC FACTOR.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO FAMILY 4 OF G-PROTEIN COUPLED RECEPTORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X91168; CAA62595.1; -.  
CC InterPro; IPR001499; -.  
CC Pfam; PF02076; STE3; 1.  
KW Transmembrane; G-protein coupled receptor; Pheromone response.  
FT TRANSMEM 8 28 POTENTIAL.  
FT TRANSMEM 36 56 POTENTIAL.  
FT TRANSMEM 70 90 POTENTIAL.  
FT TRANSMEM 110 130 POTENTIAL.  
FT NON\_TER 518 518  
SQ SEQUENCE 518 AA; 57999 MW; AA635E46A1BF7C44 CRC64;  
  
Query Match 54.4%; Score 37; DB 1; Length 518;  
Best Local Similarity 66.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 GLICPDPRP 10  
|:| | | | |  
Db 407 GVLVPDPHP 415  
  
RESULT 19  
GPV\_MOUSE  
ID GPV\_MOUSE STANDARD; PRT; 567 AA.  
AC O08742;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).  
GN GP5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Liver;  
RX MEDLINE=97275136; PubMed=9129030;  
RA Ravanat C., Morales M., Azorsa D.O., Moog S., Schuhler S.,  
RA Grunert P., Loew D., van Dorsselaer A., Cazenave J.-P., Lanza F.;  
RT "Gene cloning of rat and mouse platelet glycoprotein V:  
RT identification of megakaryocyte-specific promoters and demonstration  
RT of functional thrombin cleavage.";  
RL Blood 89:3253-3262(1997).  
CC -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND  
CC FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT  
CC PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO  
CC INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A  
CC CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z69595; CAA93441.1; -.  
CC MGD; MGI:1096363; Gp5.  
CC InterPro; IPR000483; -.  
CC InterPro; IPR001611; -.  
CC Pfam; PF00560; LRR; 13.  
CC Pfam; PF01463; LRRCT; 1.  
CC PRINTS; PR00019; LEURICHRPT.  
KW Platelet; Transmembrane; Glycoprotein; Blood coagulation;  
KW Repeat; Leucine-rich repeat; Cell adhesion; Signal.  
FT SIGNAL 1 16 POTENTIAL.  
FT CHAIN 17 567 PLATELET GLYCOPROTEIN V.  
FT DOMAIN 17 522 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 523 543 POTENTIAL.  
FT DOMAIN 544 567 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 73 96 LRR 1.  
FT REPEAT 97 120 LRR 2.  
FT REPEAT 122 144 LRR 3.  
FT REPEAT 145 168 LRR 4.  
FT REPEAT 170 192 LRR 5.  
FT REPEAT 194 216 LRR 6.  
FT REPEAT 217 240 LRR 7.  
FT REPEAT 241 264 LRR 8.  
FT REPEAT 266 288 LRR 9.  
FT REPEAT 289 312 LRR 10.  
FT REPEAT 314 337 LRR 11.  
FT REPEAT 338 361 LRR 12.  
FT REPEAT 362 385 LRR 13.  
FT REPEAT 386 409 LRR 14.  
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 567 AA; 63467 MW; C48643AA73967A7D CRC64;  
  
Query Match 54.4%; Score 37; DB 1; Length 567;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 5 CPDPRPL 11  
| | | | |  
Db 472 CPDPRSL 478  
  
RESULT 20  
LEPA\_MYCTU  
ID LEPA\_MYCTU STANDARD; PRT; 653 AA.  
AC P71739;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GTP-BINDING PROTEIN LEPA.  
GN LEPA OR RV2404C OR MTCY253.16.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,



RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
CC LEPA SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z81368; CAB03723.1; -  
CC HSSP; P13551; IELO.  
DR TubercuList; RV2404C; -  
DR InterPro; IPR000795; -  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR PROSITE; PS00301; EFATOR\_GTP; 1.  
KW GTP-binding. 59 GTP (BY SIMILARITY).  
FT NP\_BIND 124 128 GTP (BY SIMILARITY).  
FT NP\_BIND 178 181 GTP (BY SIMILARITY).  
SQ SEQUENCE 653 AA; 72395 MW; DA4AFE10E6C25755 CRC64;  
  
Query Match 54.4%; Score 37; DB 1; Length 653;  
Best Local Similarity 45.5%; Pred No 86;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 GLLCPDPRPLE 12  
Db 285 GIVSPEKPKCE 295  
|::|::|::|  
  
RESULT 21  
NRTC\_SYNP7  
ID NRTC\_SYNP7 STANDARD; PRT; 659 AA.  
AC P38045;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE NITRATE TRANSPORT ATP-BINDING PROTEIN NRTC.  
GN NRTC.  
OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
OX NCBI\_TaxID=1140;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93173091; PubMed=8437564;  
RA Omata T., Andriess X., Hirano A.;  
RT "Identification and characterization of a gene cluster involved in  
RT nitrate transport in the cyanobacterium Synecococcus sp. PCC7942.";  
RL Mol. Gen. Genet. 236:193-202(1993).  
CC -!- FUNCTION: PROBABLY PART OF A HIGH-AFFINITY BINDING-PROTEIN-  
CC DEPENDENT TRANSPORT SYSTEM FOR NITRATE. PROBABLY RESPONSIBLE FOR  
CC ENERGY COUPLING TO THE TRANSPORT SYSTEM.  
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
CC (ABC TRANSPORTERS).  
CC -!- SIMILARITY: SOME, IN THE C-TERMINAL DOMAIN TO NRTA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; X61625; CAA43811.1; -  
DR PIR; S30893; S30893.  
DR InterPro; IPR001617; -  
DR Pfam; PF00005; ABC\_tran; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
KW Transport; ATP-binding; Membrane; Nitrate assimilation.  
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
CC LEPA SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z81368; CAB03723.1; -  
CC HSSP; P13551; IELO.  
DR TubercuList; RV2404C; -  
DR InterPro; IPR000795; -  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR PROSITE; PS00301; EFATOR\_GTP; 1.  
KW GTP-binding. 59 GTP (BY SIMILARITY).  
FT NP\_BIND 124 128 GTP (BY SIMILARITY).  
FT NP\_BIND 178 181 GTP (BY SIMILARITY).  
SQ SEQUENCE 653 AA; 72395 MW; DA4AFE10E6C25755 CRC64;  
  
Query Match 54.4%; Score 37; DB 1; Length 659;  
Best Local Similarity 77.8%; Pred No 86;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AGLLCPDPR 9  
Db 506 AGELCDDPR 514  
|::|::|::|  
  
RESULT 22  
UNG\_CHLTR  
ID UNG\_CHLTR STANDARD; PRT; 229 AA.  
AC O84613;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE URACIL-DNA GLYCOSYLASE (EC 3.2.2.-) (UDG).  
GN UNG OR CT607.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=D/UW-3/CX; PubMed=9784136;  
RX MEDLINE=9900809; PubMed=9784136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis."  
RL Science 282:754-759(1998).  
CC -!- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE  
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA  
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE001331; AAC68210.1; -  
CC HSSP; P12295; IEUI.  
DR InterPro; IPR002043; -  
DR Pfam; PF00315; UNG; 1.  
DR PROSITE; PS00130; U\_DNA\_GLYCOSYLASE; 1.  
KW DNA repair; Hydrolase; Glycosidase.  
FT ACT\_SITE 70 70  
SQ SEQUENCE 229 AA; 26002 MW; 9AD49846F8DCDC3B CRC64;  
  
Query Match 52.9%; Score 36; DB 1; Length 229;  
Best Local Similarity 63.6%; Pred No 47;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPL 11  
| | | | |  
Db 187 AVLACPHPSPL 197

RESULT 23  
IBP3\_BOVIN STANDARD; PRT; 291 AA.  
AC P20959;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 3 PRECURSOR (IGFBP-3)  
GN (IBP-3) (IGF-BINDING PROTEIN 3).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91282738; PubMed=1711841;  
RA Spratt S.K., Tatsuno G.P., Sommer A.;  
RT "Cloning and characterization of bovine insulin-like growth factor  
binding protein-3 (bIGFBP-3).";  
RL Biochem. Biophys. Res. Commun. 177:1025-1032(1991).  
RN [2]  
RP SEQUENCE OF 28-52.  
RX MEDLINE=91065246; PubMed=1701128;  
RA Conover C.A., Ronk M., Lombana F., Powell D.R.;  
RT "Structural and biological characterization of bovine insulin-like  
growth factor binding protein-3.";  
RL Endocrinology 127:2795-2803(1990).  
RN [3]  
RP SEQUENCE OF 194-266 FROM N.A.  
RX MEDLINE=98030060; PubMed=9363609;  
RA Maciulla J.H., Zhang H.M., Denise S.K.;  
RT "A novel polymorphism in the bovine insulin-like growth factor  
binding protein-3 (IGFBP3) gene.";  
RL Anim. Genet. 28:375-375(1997).  
CC -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS  
AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE  
INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.  
CC -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH  
IGF-I OR IGF-II AND A 85 KDA GLYCOPROTEIN (ALS).  
CC -!- SUBCELLULAR LOCATION: SECRETED.  
CC -!- TISSUE SPECIFICITY: PLASMA; EXPRESSED BY MOST TISSUES.  
CC -!- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.  
CC -!- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
PROTEIN FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M76478; AAA30582.1; -.  
DR EMBL; U83465; AAB41430.1; -.  
DR PIR; JN0064; JN0064.  
DR InterPro; IPR000716; -.  
DR InterPro; IPR000867; -.  
DR Pfam; PF00219; IGFBP; 1.  
DR Pfam; PF00086; thyroglobulin\_1; 1.  
DR PROSITE; PS00222; IGF\_BINDING; 1.  
DR PROSITE; PS00484; THYROGLOBULIN\_1; 1.  
KW Growth factor binding; Signal; Glycoprotein.  
FT SIGNAL 1 27

FT CHAIN 28 291 INSULIN-LIKE GROWTH FACTOR BINDING  
FT PROTEIN 3.  
FT DOMAIN 28 134 IGF-BINDING (POTENTIAL).  
FT DOMAIN 138 161 SER/THR-RICH.  
FT DOMAIN 192 208 SER/THR-RICH.  
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 236 285 THYROGLOBULIN TYPE I.  
SQ SEQUENCE 291 AA; 31556 MW; E6A972535AFF241 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 291;  
Best Local Similarity 50.0%; Pred. No. 59;  
Matches 8; Conservative 2; Mismatches 2; Indels 4; Gaps 1;

QY 1 AGLLCP-----DPRPLE 12  
: | | | |  
Db 92 SGLRCQPPPGDPRPLQ 107

RESULT 24  
TCBF\_PSEQ STANDARD; PRT; 352 AA.  
AC P27101;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE MALEYLACETATE REDUCTASE (EC 1.3.1.32).  
GN TCBF.  
OS Pseudomonas sp. (strain P51).  
OG Plasmid pp51.  
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae.  
OX NCBI\_TaxID=65067;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91193197; PubMed=2013566;  
RA van der Meer J.R., Eggen R.I., Zehnder A.J., de Vos W.M.;  
RT "Sequence analysis of the Pseudomonas sp. strain P51 tcb gene  
cluster, which encodes metabolism of chlorinated catechols: evidence  
for specialization of catechol 1,2-dioxygenases for chlorinated  
substrates.";  
RL J. Bacteriol. 173:2425-2434(1991).  
CC -!- CATALYTIC ACTIVITY: 3-OXOADIPATE + NAD(P)(+) = 2-MALEYLACETATE +  
NAD(P)H.  
CC -!- PATHWAY: 3-CHLOROCATECHOL DEGRADATION (BETA-KETOADIPATE PATHWAY).  
CC THIS PATHWAY SERVES A VITAL ROLE IN THE BIODEGRADATION OF TOXIC  
AROMATIC COMPOUNDS INTRODUCED IN THE ENVIRONMENT BOTH AS NATURAL  
PRODUCTS AND AS INDUSTRIAL EFFLUENT.  
CC -!- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE  
FAMILY.  
CC -!- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A TRANS-  
DIENELACTONEISOMERASE.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M57629; AAD13629.1; -.  
DR PIR; E43673; E43673.  
DR InterPro; IPR001670; -.  
DR Pfam; PF00465; Fe-ADH; 2.  
DR PROSITE; PS00913; ADH\_IRON\_1; FALSE\_NEG.  
DR PROSITE; PS00060; ADH\_IRON\_2; FALSE\_NEG.  
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.  
SQ SEQUENCE 352 AA; 37499 MW; C72A9D2671FFFAB3 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 352;

```
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PDPRLP 12
Db 327 PNPRLP 333

RESULT 25
TFTE_BURCE
ID TFTE_BURCE STANDARD; PRT; 352 AA.
AC Q45072;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MALEYLACETATE REDUCTASE (EC 1.3.1.32).
GN TFTE
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC NCBI_TaxID=292;
OC {1}
OC SEQUENCE FROM N.A.
OC STRAIN-AC1100;
OC MEDLINE-95266809; PubMed-7538273;
OC Daubaras D.L., Hershberger C.D., Kitano K., Chakrabarty A.M.;
OC "Sequence analysis of a gene cluster involved in metabolism of 2,4,5-
OC trichlorophenoxyacetic acid by Burkholderia cepacia AC1100.";
OC Appl. Environ. Microbiol. 61:1279-1289(1995).
OC -1- CATALYTIC ACTIVITY: 3-OXOADIPATE + NAD(P)(+) - 2-MALEYLACETATE +
OC NAD(P)H.
OC -1- PATHWAY: INVOLVED IN THE DEGRADATION OF 2,4,5-
OC TRICHLOROPHENOXYACETIC ACID.
OC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
OC FAMILY.
OC -----
OC This SWISS-PROT entry is copyright. It is produced through a collaboration
OC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
OC the European Bioinformatics Institute. There are no restrictions on its
OC use by non-profit institutions as long as its content is in no way
OC modified and this statement is not removed. Usage by and for commercial
OC entities requires a license agreement (See http://www.isb-sib.ch/announce/
OC or send an email to license@isb-sib.ch).
OC -----
OC EMBL; U19883; AAC43333.1;
OC InterPro: IPR001670;
OC Pfam; PF00465; Fe-ADH; 1
OC PROSITE; PS00913; ADH_IRON_1; FALSE_NEG.
OC Aromatic hydrocarbons catabolism; Oxidoreductase; NAD.
OC SEQUENCE 352 AA; 36823 MW; FD7F0F8959357044 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 352;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PDPRLP 12
Db 328 PNPRLP 334

RESULT 26
TFFL_ALCEU
ID TFFL_ALCEU STANDARD; PRT; 354 AA.
AC P27137;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MALEYLACETATE REDUCTASE I (EC 1.3.1.32).
GN TFFL OR TFFL
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Plasmid pJP4.
```

```
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
NCBI_TaxID=510;
OC {1}
OC SEQUENCE FROM N.A.
OC STRAIN-JMP134;
OC MEDLINE-90236889; PubMed-2185214;
OC Perkins E.J., Gordon M.P., Caceres O., Lurquin P.F.;
OC "Organization and sequence analysis of the 2,4-dichlorophenol
OC hydroxylase and dichlorocatechol oxidative operons of plasmid pJP4.";
OC J. Bacteriol. 172:2351-2359(1990).
OC -1- CATALYTIC ACTIVITY: 3-OXOADIPATE + NAD(P)(+) - 2-MALEYLACETATE +
OC NAD(P)H.
OC -1- PATHWAY: 3-CHLOROCATECHOL DEGRADATION (BETA-KETOADIPATE PATHWAY).
OC THIS PATHWAY SERVES A VITAL ROLE IN THE BIODEGRADATION OF TOXIC
OC AROMATIC COMPOUNDS INTRODUCED IN THE ENVIRONMENT BOTH AS NATURAL
OC PRODUCTS AND AS INDUSTRIAL EFFLUENT.
OC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
OC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
OC FAMILY.
OC -----
OC This SWISS-PROT entry is copyright. It is produced through a collaboration
OC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
OC the European Bioinformatics Institute. There are no restrictions on its
OC use by non-profit institutions as long as its content is in no way
OC modified and this statement is not removed. Usage by and for commercial
OC entities requires a license agreement (See http://www.isb-sib.ch/announce/
OC or send an email to license@isb-sib.ch).
OC -----
OC EMBL; M35097; AAA98265.1;
OC PIR; D35255; D35255.
OC InterPro: IPR001670;
OC Pfam; PF00465; Fe-ADH; 1
OC PROSITE; PS00913; ADH_IRON_1; FALSE_NEG.
OC PROSITE; PS00060; ADH_IRON_2; FALSE_NEG.
OC Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.
OC SEQUENCE 354 AA; 37899 MW; 7D88C001B5B13376 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 354;
Best Local Similarity 85.7%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PDPRLP 12
Db 328 PNPRLP 334

RESULT 27
HXB2_HUMAN
ID HXB2_HUMAN STANDARD; PRT; 356 AA.
AC P14652; P17485; P10913;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HOMEOBOX PROTEIN HOX-B2 (HOX-2H) (HOX-2.8) (K8).
GN HOXB2 OR HOX2H.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OC {1}
OC SEQUENCE FROM N.A.
OC MEDLINE-90098876; PubMed-2574852;
OC Acampora D., D'Esposito M., Faiella A., Pannese M., Migliacchio E.,
OC Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
OC "The human HOX gene family.";
OC Nucleic Acids Res. 17:10385-10402(1989).
OC [2]
OC SEQUENCE OF 132-208 FROM N.A.
OC TISSUE-Placenta;
OC MEDLINE-89378558; PubMed-2570724;
OC Giampaolo A., Acampora D., Zappavigna V., Pannese M.,
```

RA D'Esposito M., Care A., Faiella A., Stornaiuolo A., Russo G.,  
RA Simeone A., Boncinelli E., Peschle C.;  
RT "Differential expression of human HOX-2 genes along the anterior-  
RL posterior axis in embryonic central nervous system.";  
RN Differentiation 40:191-197(1989).  
RP [3]  
RX SEQUENCE OF 143-208 FROM N.A.  
RA MEDLINE=90215256; PubMed=2576652;  
RA Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,  
RA Gaudino G., Stornaiuolo A., Cafiero M., Faiella A., Simeone A.;  
RT "Organization of human class I homeobox genes.";  
RL Genome 31:745-756(1989).  
RN [4]  
RP SEQUENCE OF 143-202 FROM N.A.  
RX MEDLINE=88329001; PubMed=2901346;  
RA Kongsuwan K., Webb E., Housiaux P., Adams J.M.;  
RT "Expression of multiple homeobox genes within diverse mammalian  
RT haemopoietic lineages.";  
RL EMBO J. 7:2131-2138(1988).  
RN [5]  
RP SEQUENCE OF 1-42 FROM N.A.  
RX MEDLINE=95181447; PubMed=7876223;  
RA Vieille-Grosjean I., Huber P.;  
RT "Transcription factor GATA-1 regulates human HOXB2 gene expression in  
RT erythroid cells.";  
RL J. Biol. Chem. 270:4544-4550(1995).  
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE EMBRYOS AND FETUSES AT  
CC 5-9 WEEKS FROM CONCEPTION.  
CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.  
CC PROBOSCIPEDIA SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X16665; CAA34655.1; -.  
DR EMBL; X16176; CAA34298.1; -.  
DR EMBL; X14571; CAA32709.1; -.  
DR EMBL; X78978; CAA55581.1; -.  
DR PIR; S07542; WJHU2H.  
DR PIR; E37042; E37042.  
DR HSSP; P02833; 1SAN.  
DR MIM; 142967; -.  
DR InterPro; IPR001356; -.  
DR InterPro; IPR001827; -.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEOBOX.  
DR PRINTS; PR00025; ANTENNAPEDIA.  
DR PROSITE; PS00027; HOMEOBOX\_1; 1.  
DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
DR PROSITE; PS50071; HOMEOBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 94 99 ANTP-TYPE HEXAPEPTIDE.  
FT DNA\_BIND 143 202 HOMEOBOX.  
FT CONFLICT 136 138 PEA -> RRL (IN REF. 2).  
SQ SEQUENCE 356 AA; 37913 MW; D7ACA0E3D5BEBFB9 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 356;  
Best Local Similarity 70.0%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 GLLCPDRPL 11  
| | |||||

Db 254 GALSADPRPL 263  
RESULT 28  
CA26\_CHICK  
ID CA26\_CHICK STANDARD; PRT; 1022 AA.  
AC P15988;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE COLLAGEN ALPHA 2(VI) CHAIN PRECURSOR.  
GN COL6A2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89305506; PubMed=2787244;  
RA Koller E., Winterhalter K.H., Trueb B.;  
RT "The globular domains of type VI collagen are related to the  
RT collagen-binding domains of cartilage matrix protein and von  
RT Willebrand factor.";  
RL EMBO J. 8:1073-1077(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91200044; PubMed=2015818;  
RA Hayman A.R., Koppel J., Trueb B.;  
RT "Complete structure of the chicken alpha 2(VI) collagen gene.";  
RL Eur. J. Biochem. 197:177-184(1991).  
RN [3]  
RP SEQUENCE OF 8-38 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=91187664; PubMed=2011522;  
RA Koller E., Hayman A.R., Trueb B.;  
RT "The promoter of the chicken alpha 2(VI) collagen gene has features  
RT characteristic of house-keeping genes and of proto-oncogenes.";  
RL Nucleic Acids Res. 19:485-491(1991).  
CC -!- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.  
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),  
CC ALPHA 2(VI), AND ALPHA 3(VI).  
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X15041; CAA33144.1; -.  
DR EMBL; X56659; CAA39982.1; -.  
DR EMBL; X56595; CAA39933.1; -.  
DR PIR; S04111; S04111.  
DR PIR; S23378; S23378.  
DR InterPro; IPR000087; -.  
DR InterPro; IPR002035; -.  
DR Pfam; PF01391; Collagen; 5.  
DR Pfam; PF00092; vwa; 3.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR PROSITE; PS50234; VWFA\_DOMAIN; 3.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Cell adhesion; Collagen; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 1022 COLLAGEN ALPHA 2(VI) CHAIN.  
FT DOMAIN 28 255 NONHELICAL REGION.  
FT DOMAIN 256 590 TRIPLE-HELICAL REGION.  
FT DOMAIN 591 1022 NONHELICAL REGION.  
FT SITE 514 519 INTERUPTION IN COLLAGENOUS REGION.  
FT DOMAIN 44 168 VWFA 1.



FT DOMAIN 613 738 VF6A 2. 738  
FT SITE 833 957 VF6A 3.  
FT SITE 348 350 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 366 368 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 426 428 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 444 446 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 465 467 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 489 491 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 498 500 CELL ATTACHMENT SITE (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 897 897 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1022 AA; 109176 MW; 5194CFD1475AE893 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 1022;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 LLCDDPR 9  
DB 804 MLCDDPO 810

RESULT 29  
UTY\_MOUSE  
ID UTY\_MOUSE STANDARD; PRT; 1212 AA.  
AC P79457: 097979;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE UBQUITOUSLY TRANSCRIBED Y CHROMOSOME TETRATRICOPEPTIDE REPEAT PROTEIN  
DE (UBQUITOUSLY TRANSCRIBED Y CHROMOSOME TETRATRICOPEPTIDE REPEAT PROTEIN) (MALE-  
DE SPECIFIC HISTOCOMPATIBILITY ANTIGEN H-YDB).  
GN UTY.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R11; TISSUE-Testis;  
RX MEDLINE=98409500; PubMed=9736773;  
RA Mazeyrat S., Saut N., Sargent C.A., Grimmond S., Longepied G., Mitchell M.J.;  
RA Ehrmann I.E., Ellis P.S., Greenfield A., Affara N.A., Mitchell M.J.;  
RT "The mouse Y chromosome interval necessary for spermatogonial  
RT proliferation is gene dense with syntenic homology to the human AzFa  
RT region.";  
RL Hum. Mol. Genet. 7:1713-1724(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SWISS;  
RX MEDLINE=97099462; PubMed=8944031;  
RA Greenfield A., Scott P., Pennisi D., Ehrmann I., Ellis P., Cooper L.,  
RA Simpson E., Koopman P.;  
RT "An H-Ydb epitope is encoded by a novel mouse Y chromosome gene.";  
RL Nat. Genet. 14:474-478(1996).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -!- SIMILARITY: CONTAINS 6 TPR REPEATS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF057367; AAC67385.1; -  
CC EMBL; Y09222; CAA70422.1; -  
CC MGD; MGI:894810; Uty.

DR InterPro; IPR001440; -  
DR Pfam; PF00515; TPR; 6.  
KW Repeat; TPR repeat; Nuclear protein.  
FT DOMAIN 91 119  
FT SITE 128 156 TPR 1.  
FT SITE 165 193 TPR 2.  
FT SITE 203 231 TPR 3.  
FT SITE 231 267 TPR 4.  
FT SITE 316 344 TPR 5.  
FT SITE 350 378 TPR 6.  
FT CONFLICT 1069 1069 E -> Q (IN REF. 2).  
FT CONFLICT 1149 1212 EVENLFFVTNENESQKTYIVHQCNCARKTSGNLENFVVLEQ  
FT YKMDLIDQVYDQFTLAPLSAS -> STRDLPLQLHLRQC  
FT HLOGPTDKAAILEFHLTEGSGDMH (IN REF. 2).  
SQ SEQUENCE 1212 AA; 136736 MW; 2AE1A816F3D6ACB5 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 1212;  
Best Local Similarity 71.4%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 LCPDPRP 10  
DB 729 ICPPRP 735

RESULT 30  
YDM5\_SCHPO  
ID YDM5\_SCHPO STANDARD; PRT; 1337 AA.  
AC P87136;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 150.8 KDA PROTEIN C57A7.05 IN CHROMOSOME I.  
GN SPAC57A7.05.  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA Skelton J., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;  
RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -!- SIMILARITY: SOME, TO YEAST YDL231C.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z95396; CAB08763.1; -  
CC Hypothetical protein; Transmembrane.  
KW TRANSMEM 209 229 POTENTIAL.  
FT TRANSMEM 241 261 POTENTIAL.  
FT TRANSMEM 267 287 POTENTIAL.  
FT TRANSMEM 328 348 POTENTIAL.  
FT TRANSMEM 361 381 POTENTIAL.  
FT TRANSMEM 387 407 POTENTIAL.  
FT TRANSMEM 917 937 POTENTIAL.  
FT TRANSMEM 975 995 POTENTIAL.  
FT TRANSMEM 997 1017 POTENTIAL.  
FT TRANSMEM 1021 1041 POTENTIAL.  
FT TRANSMEM 1066 1086 POTENTIAL.  
FT TRANSMEM 1275 1295 POTENTIAL.  
SQ SEQUENCE 1337 AA; 150848 MW; A549BC8E00D08791 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 1337;

```
Best Local Similarity 71.4%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLLCPDP 8
Db 304 GFVCPDP 310

RESULT 31
BAI2_HUMAN STANDARD; PRT; 1572 AA.
AC O60241;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 2 PRECURSOR.
GN BAI2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98194217; PubMed=9533023;
RA Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;
RT "Cloning and characterization of BAI2 and BAI3, novel genes homologous
RL to brain-specific angiogenesis inhibitor 1 (BAI1).";
RL Cytogenet. Cell Genet. 79:103-108(1997).
CC -!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN
CC HEART, THYMUS, SKELETAL MUSCLE, AND DIFFERENT CELL LINES.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -!- SIMILARITY: CONTAINS 4 TYPE-1 TSP DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB005298; BAA25362.1; -.
DR MIM; 602683; -.
DR InterPro; IPR000203; -.
DR InterPro; IPR000832; -.
DR InterPro; IPR000884; -.
DR InterPro; IPR001879; -.
DR Pfam; PF00002; 7tm2; 1.
DR Pfam; PF00090; tsp_1; 4.
DR Pfam; PF01825; GPS; 1.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Repeat.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1572 BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 2.
FT DOMAIN 21 924 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 925 945 1 (POTENTIAL).
FT DOMAIN 946 953 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 954 974 2 (POTENTIAL).
FT DOMAIN 975 982 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 983 1003 3 (POTENTIAL).
FT DOMAIN 1004 1024 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1025 1045 4 (POTENTIAL).
FT DOMAIN 1046 1066 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1067 1087 5 (POTENTIAL).
FT DOMAIN 1088 1141 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1142 1162 6 (POTENTIAL).
FT DOMAIN 1163 1168 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1169 1189 7 (POTENTIAL).
```

```
FT DOMAIN 1190 1572 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 297 351 TSP TYPE-1 1.
FT DOMAIN 352 406 TSP TYPE-1 2.
FT DOMAIN 407 461 TSP TYPE-1 3.
FT DOMAIN 463 517 TSP TYPE-1 4.
FT DOMAIN 117 122 POLY-GLU.
FT DOMAIN 177 180 POLY-ASN.
FT DOMAIN 222 225 POLY-THR.
FT DOMAIN 1303 1306 POLY-PRO.
FT DOMAIN 1352 1358 POLY-GLY.
FT DOMAIN 1413 1418 POLY-PRO.
FT CARBOHYD 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 855 855 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1572 AA; 171140 MW; A9775645B77BC285 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 1572;
Best Local Similarity 66.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDP 10
Db 813 GLILPPRP 821

RESULT 32
Y192_HUMAN STANDARD; PRT; 2124 AA.
AC Q93074;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0192 (FRAGMENT).
GN KIAA0192.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D83783; BAA12112.1; -.
KW Hypothetical protein.
FT NON_TER 1 1
FT DOMAIN 599 602 POLY-SER.
FT DOMAIN 1201 1207 POLY-GLY.
FT DOMAIN 1998 2124 GLN-RICH.
FT DOMAIN 1998 2023 POLY-GLN.
FT DOMAIN 2028 2033 POLY-GLN.
FT DOMAIN 2037 2070 POLY-GLN.
FT DOMAIN 2090 2097 POLY-GLN.
```

```
SQ SEQUENCE 2124 AA; 237207 MW; 255FB9419EC39F42 CRC64;
Query Match 52.9%; Score 36; DB 1; Length 2124;
Best Local Similarity 66.7%; Pred. NO. 3.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11
DB 297 LMCPOHRPL 305
|:|:|:|:|:|
|:|:|:|:|:|

RESULT 33
RU2A_ARATH
ID P43333; 004497; STANDARD; PRT; 249 AA.
AC 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE U2 SMALL NUCLEAR RIBONUCLEOPROTEIN A' (U2 SNRNP-A').
GN AT1G09760 OR F21M12.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24; TISSUE=Flower buds;
RA Yu D.Y., Quigley F., Mache R.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA; PubMed=11130712;
RX MEDLINE=21016719;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Egtu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziali A.,
RA Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pail G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RA thaliana.";
RL Nature 408:816-820(2000).
[3]
RP SEQUENCE OF 1-110 FROM N.A.
RC STRAIN=CV. C24; TISSUE=Flower buds;
RA Mache R., Quigley F., Thomas F., Yu D.Y.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
[4]
RP PROTEIN TO BIND STEM LOOP IV OF U2 SNRNP (BY SIMILARITY).
CC -1- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH SN-RNP U2. IT HELPS THE
CC A' PROTEIN TO BIND STEM LOOP IV OF U2 SNRNP (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X69137; CAA48890.1; -;
CC Query Match 52.2%; Score 35.5; DB 1; Length 501;
CC Best Local Similarity 63.6%; Pred. No. 1.2e+02;
CC -----
DR EMBL; AC000132; AAB60746.1; -;
DR EMBL; Z18200; CAA79133.1; -;
DR Mendel; 17411; Arath;1851;17411.
DR InterPro; IPR001611; -;
DR Pfam; PF00560; LRR; 2.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Leucine-rich repeat;
KW Repeat.
FT REPEAT 16 41 LRR 1.
FT REPEAT 42 63 LRR 2.
FT REPEAT 64 86 LRR 3.
FT REPEAT 88 111 LRR 4.
FT REPEAT 112 136 LRR 5.
FT CONFLICT 107 107 D -> H (IN REF. 1 AND 3).
FT CONFLICT 129 129 A -> P (IN REF. 1).
FT SEQUENCE 249 AA; 28041 MW; F51CB18FD86DD2 CRC64;
SQ SEQUENCE 249 AA; 28041 MW; F51CB18FD86DD2 CRC64;

Query Match 52.2%; Score 35.5; DB 1; Length 249;
Best Local Similarity 47.1%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 2; Indels 5; Gaps 1;

QY 1 AGLCCPDP-----RPLE 12
DB 232 AGLIIPDPATNDSAPME 248
|:|:|:|:|:|
|:|:|:|:|:|

RESULT 34
GSPE_AERHY
ID GSPE_AERHY STANDARD; PRT; 501 AA.
AC P31741;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN E (TYPE II TRAFFIC WARDEN ATPASE).
GN EXEE.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=644;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AH65;
RX MEDLINE=92349963; PubMed=1640836;
RA Jiang B., Howard S.P.;
RA "The Aeromonas hydrophila exee gene, required both for protein
RA secretion and normal outer membrane biogenesis, is a member of a
RA general secretion pathway.";
RL Mol. Microbiol. 6:1351-1361(1992).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE PULE/OUTE/EXEE/XPSE/XCPR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66504; CAA47126.1; -;
CC PIR; S22669; S22669.
CC InterPro; IPR001482; -;
CC Pfam; PF00437; GSP1_E; 1.
CC PROSITE; PS00662; T2SP_E; 1.
CC Transport; ATP-binding.
FT NP_BIND 262 269 ATP (POTENTIAL).
FT SEQUENCE 501 AA; 55882 MW; D8F549B80AA826DE CRC64;
SQ SEQUENCE 501 AA; 55882 MW; D8F549B80AA826DE CRC64;

Query Match 52.2%; Score 35.5; DB 1; Length 501;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
CC -----
DR EMBL; X69137; CAA48890.1; -;
```

Matches 7; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 4 LCPD---PRPL 11  
|||||  
Db 394 LCPDCRAPRPI 404

RESULT 35  
PLF4\_SHEEP STANDARD; PRT; 85 AA.  
AC P30035;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PLATELET FACTOR 4 (PF-4).  
GN SCYB4 OR PF4.

OS Eukaryotes (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]

RP SEQUENCE.  
RX MEDLINE=92160127; PubMed=1788836;  
RA Shigeta O., Lu W., Holt J.C., Edmunds L.H. Jr., Niewiarowski S.;  
RT "Ovine platelet factor 4: purification, amino acid sequence,  
RT radioimmunoassay and comparison with platelet factor 4 of other  
RT species.";  
RL Thromb. Res. 64:509-520(1991).

CC -1- FUNCTION: PLATELET FACTOR 4, NONCOVALENTLY BOUND TO A PROTEOGLYCAN  
CC MOLECULE, IS RELEASED DURING PLATELET AGGREGATION. PF4 NEUTRALIZES  
CC THE ANTICOAGULANT EFFECT OF HEPARIN BECAUSE IT BINDS MORE STRONGLY  
CC TO HEPARIN THAN TO THE CHONDROITIN-4-SULFATE CHAINS OF THE CARRIER  
CC MOLECULE. CHEMOTACTIC FOR NEUTROPHILS AND MONOCYTES.

CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE  
CC C-X-C) (CHEMOKINE CXC).

DR HSSP; P02777; 1PLF.  
DR InterPro; IPR001089; -.  
DR InterPro; IPR001811; -.  
DR Pfam; PF00048; IL8; 1.  
DR PRINTS; PR00437; SMALLCYTKCXC.  
DR PROSITE; PS00471; SMALL\_CYTOKINES\_CXC; 1.  
KW Cytokine; Platelet; Proteoglycan; Heparin-binding; Chemotaxis.  
FT DISULFID 25 51 BY SIMILARITY.  
FT DISULFID 27 67 BY SIMILARITY.  
SQ SEQUENCE 85 AA; 9129 MW; 925A76512095ECF7 CRC64;

Query Match 51.5%; Score 35; DB 1; Length 85;  
Best Local Similarity 66.7%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLLCPDPR 9  
|||||  
Db 47 AGLHCPSPQ 55

RESULT 36  
SR19\_ORYSA STANDARD; PRT; 136 AA.  
AC P49964;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE SIGNAL RECOGNITION PARTICLE 19 KDA PROTEIN (SRP19).  
GN SRP19.

OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;  
OC Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE / JAPONICA; TISSUE=Callus;  
RA Zwieb C.W., Black S.D.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY, BINDS DIRECTLY  
CC TO 7S RNA AND MEDIATES BINDING OF THE 54 KDA SUBUNIT OF THE SRP.  
CC (BY SIMILARITY).  
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE  
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,  
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE SRP19 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
DR EMBL; U19030; AAB65810.1; -.  
DR InterPro; IPR002778; -.  
DR Pfam; PF01922; SRP19; 1.  
KW Signal recognition particle; RNA-binding; Ribonucleoprotein.  
FT DOMAIN 127 136 BASIC REGION, POTENTIALLY INVOLVED IN  
FT RNA-BINDING.  
SQ SEQUENCE 136 AA; 14922 MW; 65D3B91967F3D638 CRC64;

Query Match 51.5%; Score 35; DB 1; Length 136;  
Best Local Similarity 50.0%; Pred. No. 43;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPLE 12  
:| |||| :|  
Db 35 SGKACPDPTCVE 46

RESULT 37  
IBP3\_PIG STANDARD; PRT; 266 AA.  
ID IBP3\_PIG  
AC P16611;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 3 (IGFBP-3) (IBP-3) (IGF-  
DE BINDING PROTEIN 3).  
GN IGFBP3.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90130475; PubMed=1688850;  
RA Shimazaki S., Shimonaka M., Ui M., Inouye S., Shibata F., Ling N.;  
RT "Structural characterization of a follicle-stimulating hormone action  
RT inhibitor in porcine ovarian follicular fluid. Its identification as  
RT the insulin-like growth factor-binding protein.";  
RL J. Biol. Chem. 265:2198-2202(1990).  
RN [2]

RP SEQUENCE OF 1-15.  
RX MEDLINE=92109718; PubMed=1722398;  
RA Coleman M.E., Pan Y.-C.E., Ethernan T.D.;  
RT "Identification and NH2-terminal amino acid sequence of three  
RT insulin-like growth factor-binding proteins in porcine serum.";  
RL Biochem. Biophys. Res. Commun. 181:1131-1136(1991).  
CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS  
CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
CC PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE  
CC INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.  
CC -1- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH





CC -----  
DR EMBL; X58778; CAA41578.1; -.  
DR PIR; S20452; S20452.  
DR MEROPS; M19.003; -.  
DR InterPro; IPR000180; -.  
DR PROSITE; PS00869; RENAL\_Dipeptidase; 1.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 271 AA; 29484 MW; CB8AB11F3C8EC42E CRC64;  
  
Query Match 51.5%; Score 35; DB 1; Length 271;  
Best Local Similarity 63.6%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 AGLICPDPRPL 11  
Db 150 AHALCPQPRNL 160  
  
RESULT 40  
FCN3\_HUMAN  
ID FCN3\_HUMAN STANDARD; PRT; 299 AA.  
AC 075636;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE FICOLIN 3 PRECURSOR (COLLAGEN/FIBRINOGEN DOMAIN-CONTAINING PROTEIN 3)  
DE (COLLAGEN/FIBRINOGEN DOMAIN-CONTAINING LECTIN 3 P35) (HAKATA ANTIGEN).  
GN FCN3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Lung;  
RX MEDLINE=98361935; PubMed=9694814;  
RA Sugimoto R., Yae Y., Akaiwa M., Kitajima S., Shibata Y., Sato H.,  
RA Hirata J., Okochi K., Izuhara K., Hamasaki N.;  
RT "Cloning and characterization of the Hakata antigen, a member of the  
RT ficolin/opsolin p35 lectin family.";  
RL J. Biol. Chem. 273:20721-20727(1998).  
RN [2]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=99264375; PubMed=10330454;  
RA Akaiwa M., Yae Y., Sugimoto R., Suzuki S.O., Iwaki T., Izuhara K.,  
RA Hamasaki N.;  
RT "Hakata antigen, a new member of the ficolin/opsolin p35 family, is a  
RT novel human lectin secreted into bronchus/alveolus and bile.";  
RL J. Histochem. Cytochem. 47:777-786(1999).  
CC -!- FUNCTION: INVOLVED IN THE SERUM EXERTING LECTIN ACTIVITY. HAS  
CC AFFINITY WITH GALNAC, GLCNAC, D-FUCOSE, AS MONO/OLIGOSACCHARIDE  
CC AND LIPOPOLYSACCHARIDES FROM S.TYPHIMURIUM AND S.MINNESOTA.  
CC -!- SUBUNIT: DISULFIDE-LINKED HOMOPOLYMER. MAY BE A OCTADECAMER  
CC CONSISTING OF AN ELEMENTARY TRIMER UNIT. DOES NOT INTERACT WITH  
CC FIBRONECTIN, ELASTIN OR ZYMOSAN.  
CC -!- SUBCELLULAR LOCATION: SECRETED IN PLASMA, BRONCHUS, ALVEOLUS AND  
CC BILE DUCT.  
CC -!- TISSUE SPECIFICITY: LIVER AND LUNG. IN LIVER IT IS PRODUCED BY  
CC BILE DUCT EPITHELIAL CELLS AND HEPATOCYTES. IN LUNG IT IS PRODUCED  
CC BY BOTH CILIATED BRONCHIAL EPITHELIAL CELLS AND TYPE II ALVEOLAR  
CC EPITHELIAL CELLS.  
CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
CC -!- DISEASE: ANTIGEN FOUND 14 TIMES MORE FREQUENTLY IN THE SERUM OF  
CC PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS (SLE) THAN IN PATIENTS  
CC WITH OTHER AUTOIMMUNE DISEASES.  
CC -!- SIMILARITY: BELONGS TO THE FICOLIN LECTIN FAMILY.  
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D88587; BAA32277.1; -.  
DR MIM; 604973; -.  
DR HSSP; P02671; 1FZD.  
DR InterPro; IPR000087; -.  
DR InterPro; IPR002181; -.  
DR Pfam; PF00147; fibrinogen\_C; 1.  
DR PROSITE; PS00514; FIBRIN\_AG\_C\_DOMAIN; 1.  
KW Lectin; Collagen; Repeat; Glycoprotein; Signal; Multigene family;  
KW Antigen; Hydroxylation.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 299 FICOLIN 3.  
FT DOMAIN 48 80 COLLAGEN-LIKE.  
FT DOMAIN 119 265 FIBRINOGEN C-TERMINAL.  
FT MOD\_RES 50 50 HYDROXYLATION.  
FT MOD\_RES 53 53 HYDROXYLATION.  
FT MOD\_RES 59 59 HYDROXYLATION.  
FT MOD\_RES 65 65 HYDROXYLATION.  
FT MOD\_RES 68 68 HYDROXYLATION.  
FT MOD\_RES 77 77 HYDROXYLATION.  
FT CARBOHYD 189 189 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 299 AA; 32889 MW; 5CB8A7D3679FB264 CRC64;  
  
Query Match 51.5%; Score 35; DB 1; Length 299;  
Best Local Similarity 75.0%; Pred. No. 88;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 5 CPDPRPLE 12  
Db 29 CPGPRELE 36  
  
RESULT 41  
RM04\_YEAST  
ID RM04\_YEAST STANDARD; PRT; 319 AA.  
AC P36517;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE 60S RIBOSOMAL PROTEIN L4, MITOCHONDRIAL PRECURSOR (YML4).  
GN MRPL4 OR YLR439W OR L9753.1.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=07173;  
RA Graack H.-R.;  
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,  
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
RA Rifkin L., Riles L., Taich A., Trevaskis E., Vignati D.,  
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 15-41.  
RX MEDLINE=89078618; PubMed=3060376;  
RA Graack H.-R., Grohmann L., Choli T.;  
RT "Mitochondrial ribosomes of yeast: isolation of individual proteins  
RT and N-terminal sequencing.";



RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF232689; AAC58815.1; -.  
DR GCRDb; GCR\_1283; -.  
DR InterPro; IPR000276; -.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; 1.  
DR PROSITE; PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
KW G-protein coupled receptor; Transmembrane.  
FT DOMAIN 1 33 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 34 61 1 (POTENTIAL).  
FT DOMAIN 62 71 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 72 94 2 (POTENTIAL).  
FT DOMAIN 95 107 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 108 129 3 (POTENTIAL).  
FT DOMAIN 130 150 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 151 169 4 (POTENTIAL).  
FT DOMAIN 170 204 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 205 224 5 (POTENTIAL).  
FT DOMAIN 225 244 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 245 268 6 (POTENTIAL).  
FT DOMAIN 269 285 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 286 309 7 (POTENTIAL).  
FT DOMAIN 310 387 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 375 385 POLY-PRO.  
FT CARBOHYD 20 20 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 387 AA; 43159 MW; EF80D3F10344D6AE CRC64;

Query Match 51.5%; Score 35; DB 1; Length 387;  
Best Local Similarity 54.5%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPL 11  
| : | | |  
Db 159 ASLMCASAPL 169

RESULT 45  
GLNA\_HALVO  
ID GLNA\_HALVO STANDARD; PRT; 454 AA.  
AC P43386;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GLUTAMINE SYNTHETASE (EC 6.3.1.2) (GLUTAMATE--AMMONIA LIGASE) (GS).  
GN GLNA.  
OS Halobacterium volcanii (Haloferax volcanii).  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.  
OX NCBI\_TaxID=2246;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DS2;  
RX MEDLINE=94365840; PubMed=7916055;  
RA Brown J.R., Masuchi Y., Robb F.T., Doolittle W.F.;  
RT "Evolutionary relationships of bacterial and archaeal glutamine  
RT synthetase genes."  
RL J. MOL. EVOL. 38:566-576(1994).  
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + GLUTAMINE +  
CC ORTHOPHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U03029; AAC43489.1; -.  
DR HSSP; P06201; 2LGS.  
DR InterPro; IPR001691; -.  
DR Pfam; PF00120; gln-synt; 1.  
DR PROSITE; PS00180; GLNA\_1; 1.  
DR PROSITE; PS00181; GLNA\_ATP; 1.  
KW Ligase.  
FT BINDING 384 384 AMP (UNDER CONDITIONS OF ABUNDANT  
FT SEQUENCE 454 AA; 50696 MW; F07ECC3133B07722 CRC64;  
SQ

Query Match 51.5%; Score 35; DB 1; Length 454;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLLCPDP 8  
| | | | |  
Db 372 GLDCPDP 378

Search completed: June 28, 2001, 11:57:25  
Job time: 266 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:30 ; Search time 72.61 seconds  
(without alignments)  
21.866 Million cell updates/sec

Title: US-09-439-313-558  
Perfect score: 68  
Sequence: 1 AGLCPDPRPLE 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SPTREMBL\_16: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_unclassified: \*  
13: sp\_vertebrate: \*  
14: sp\_virus: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	67.6	782	4 Q9UID3	Q9uid3 homo sapien
2	44	64.7	205	4 Q9Y4U5	Q9y4u5 homo sapien
3	44	64.7	757	5 Q22574	Q22574 caenorhabdi
4	43	63.2	362	13 Q92074	Q92074 g beta-1,4-
5	41	60.3	271	4 Q9HBS0	Q9hbs0 homo sapien
6	41	60.3	362	2 P95188	P95188 mycobacteri
7	41	60.3	397	2 O53870	O53870 mycobacteri
8	41	60.3	430	2 O53854	O53854 mycobacteri
9	41	60.3	568	14 Q69321	Q69321 marek's dis
10	41	60.3	574	6 O62759	O62759 sus scrofa
11	41	60.3	711	11 Q9ESJ4	Q9esj4 mus musculu
12	41	60.3	729	4 Q9UJ93	Q9uj93 homo sapien
13	41	60.3	1952	4 Q9UJ92	Q9uj92 homo sapien
14	41	60.3	2135	4 O43157	O43157 homo sapien
15	41	60.3	2135	4 Q9UIV7	Q9uiv7 homo sapien
16	40	58.8	140	2 Q9I5Q7	Q9i5q7 pseudomonas
17	40	58.8	210	10 Q9SIY1	Q9siyl arabidopsis
18	40	58.8	261	2 O86349	O86349 mycobacteri
19	40	58.8	328	14 Q68387	Q68387 human cytom

20	40	58.8	355	2 Q9L1A1	Q9llal streptomyce
21	40	58.8	522	2 Q9RNX2	Q9rnx2 neisseria g
22	40	58.8	522	2 Q9RNX1	Q9rnx1 neisseria g
23	40	58.8	522	2 Q9JY63	Q9jy63 neisseria m
24	40	58.8	522	2 Q9JT47	Q9jt47 neisseria m
25	40	58.8	528	10 O49602	O49602 arabidopsis
26	40	58.8	538	10 O80480	O80480 arabidopsis
27	40	58.8	633	2 Q9RRT1	Q9rtl1 deinococcus
28	40	58.8	1301	5 Q9U122	Q9u122 leishmania
29	40	58.8	1624	5 Q9W437	Q9w437 drosophila
30	39.5	58.1	399	10 P93810	P93810 arabidopsis
31	39	57.4	146	11 Q63552	Q63552 rattus norv
32	39	57.4	262	5 Q9V3K5	Q9v3k5 drosophila
33	39	57.4	305	11 Q9JHJ7	Q9jhj7 mus musculu
34	39	57.4	392	4 O60393	O60393 homo sapien
35	39	57.4	432	11 P70225	P70225 mus musculu
36	39	57.4	432	11 Q64385	Q64385 mus musculu
37	39	57.4	532	10 O49599	O49599 arabidopsis
38	39	57.4	532	10 O81520	O81520 arabidopsis
39	39	57.4	534	10 Q9SLX0	Q9slx0 oryza sativ
40	39	57.4	535	10 O49600	O49600 arabidopsis
41	39	57.4	740	4 Q9P200	Q9p200 homo sapien
42	39	57.4	974	5 P91658	P91658 drosophila
43	39	57.4	1124	5 Q9VYR4	Q9vyr4 drosophila
44	39	57.4	1529	3 Q9Y7C6	Q9y7c6 aspergillus
45	39	57.4	1734	11 Q9JIO9	Q9jiq9 rattus norv
46	39	57.4	1793	11 Q9JIR0	Q9jir0 rattus norv
47	38.5	56.6	529	2 Q9ZBJ5	Q9zbj5 streptomyce
48	38	55.9	190	13 Q91649	Q91649 xenopus lae
49	38	55.9	277	2 Q9WXG7	Q9wxg7 alcaligenes
50	38	55.9	338	2 Q9X1L7	Q9xl17 thermotoga
51	38	55.9	350	2 Q9RYE5	Q9rye5 deinococcus
52	38	55.9	415	5 Q9GRW0	Q9grw0 holotrichia
53	38	55.9	425	10 Q9FWI4	Q9fwi4 oryza sativ
54	38	55.9	475	5 Q9UAT5	Q9uat5 caenorhabdi
55	38	55.9	483	2 P72774	P72774 synecocyst
56	38	55.9	578	2 P95194	P95194 mycobacteri
57	38	55.9	625	2 Q9R9E7	Q9r9e7 pseudomonas
58	38	55.9	661	3 O13444	O13444 cladosporiu
59	38	55.9	883	2 Q55102	Q55102 streptomyce
60	38	55.9	1088	11 Q9R189	Q9r189 rattus norv
61	38	55.9	1090	10 Q9FS08	Q9fs08 oryza sativ
62	38	55.9	1283	4 O95451	O95451 homo sapien
63	38	55.9	1321	4 O95291	O95291 homo sapien
64	38	55.9	1566	5 Q9W2F2	Q9w2f2 drosophila
65	38	55.9	1594	4 Q9HC84	Q9hc84 homo sapien

ALIGNMENTS

RESULT	1
Q9UID3	Q9UID3
ID	Q9UID3
AC	Q9UID3; PRELIMINARY; PRT; 782 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE	ANG2.
GN	ANG2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=98277456; PubMed=9615229;
RA	Lemmens I.H., Kas K., Merregaert J., Van de Ven W.J.;
RT	"Identification and molecular characterization of TM7SF2 in the FAUNA
RL	gene cluster on human chromosome 11q13.";
RN	Genomics 49:437-442(1998).
RP	SEQUENCE FROM N.A.

RA Lemmens I.H., Kas K., Merregaert J., Van de Ven W.J.M.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF024631; AAF21627.2; -- 0E858672E4C656DD CRC64;  
SQ SEQUENCE 782 AA; 86041 MW; 0E858672E4C656DD CRC64;

Query Match 67.6%; Score 46; DB 4; Length 782;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPLE 12  
Db 760 AALRCPDVPFME 771

RESULT 2  
QY4U5 PRELIMINARY; PRT; 205 AA.

AC Q9Y4U5;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)  
DE HYPOTHETICAL 21.5 KDA PROTEIN (FRAGMENT).  
CN DKFZP586I111.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=UTERUS;  
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL050131; CAB43283.2; --

KW Hypothetical protein.  
FT NON\_TER  
SQ SEQUENCE 205 AA; 21548 MW; D88EB7A823C618E4 CRC64;

Query Match 64.7%; Score 44; DB 4; Length 205;  
Best Local Similarity 87.5%; Pred. No. 7.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CPDPRPLE 12  
Db 68 CPQPRPLE 75

RESULT 3

Q22574 PRELIMINARY; PRT; 757 AA.

AC Q22574;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE COSMID T19D12.  
GN T19D12.6.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;  
RA Willson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smalton N., Smith A., Sonnenhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.

RA Favello T.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.

RA Waterston R.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U41263; AAC24426.1; --

DR InterPro; IPR000561; --  
DR InterPro; IPR001791; --  
DR Pfam; PF00054; laminin\_G; 3.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR SMART; SM00282; LamG\_1;  
SQ SEQUENCE 757 AA; 84392 MW; 7258201954EDF997 CRC64;

Query Match 64.7%; Score 44; DB 5; Length 757;  
Best Local Similarity 58.3%; Pred. No. 24;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPLE 12  
Db 25 SGVSCPDMPPLQ 36

RESULT 4

Q92074 PRELIMINARY; PRT; 362 AA.

AC Q92074;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE BETA-1,4-GALACTOSYLTRANSFERASE (EC 2.4.1.38) (BETA-N-  
DE ACETYLGLUCOSAMINYL-GLYCOPETIDE BETA-1,4-GALACTOSYLTRANSFERASE)  
DE (GLYCOPROTEIN 4-BETA-GALACTOSYLTRANSFERASE) (THYROID  
DE GALACTOSYLTRANSFERASE) (UDP-GALACTOSE-GLYCOPROTEIN  
DE GALACTOSYLTRANSFERASE).  
GN CKI.

OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;

RN [1]  
RP SEQUENCE FROM N.A.

RA Shaper J.H.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + N-ACETYL-BETA-D-  
CC GLUCOSAMINYLGLYCOPEPTIDE = UDP + BETA-D-GALACTOSYL-1,4-N-ACETYL-  
CC BETA-D-GLUCOSAMINYLGLYCOPEPTIDE.  
DR EMBL; U19890; AAB05218.1; --

KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 362 AA; 40928 MW; E481063638EBBC9A CRC64;

Query Match 63.2%; Score 43; DB 13; Length 362;  
Best Local Similarity 87.5%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11  
Db 92 LCPDPSPL 99

RESULT 5

Q9HBS0 PRELIMINARY; PRT; 271 AA.  
ID Q9HBS0  
AC Q9HBS0;







DR EMBL; AF130313; AAF36503.1; -.  
SQ SEQUENCE 711 AA; 78706 MW; FE700EA466A251BD CRC64;

Query Match 60.3%; Score 41; DB 11; Length 711;  
Best Local Similarity 77.8%; Pred. No. 71;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11  
| | | | | | | |  
Db 246 LLCPSPL 254

RESULT 12  
Q9UJ93  
ID Q9UJ93 PRELIMINARY; PRT; 729 AA.  
AC Q9UJ93;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PLEXIN-B1/SEP RECEPTOR PRECURSOR.  
GN PLEXIN-B1/SEP.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EPITHELIUM;

RX MEDLINE=99449305; PubMed=10520995;

RA Tamagnone L., Artigiani S., Chen H., He Z., Ming G.L., Song H.L.,  
Chedotal A., Winberg M.L., Goodman C.S., Poo M.M., Tessier-Lavigne M.,  
Comoglio P.M.;

RT "Plexins are a large family of receptors for transmembrane, secreted  
and GPI-anchored semaphorins in vertebrates.";

RL Cell 99:71-80(1999).

DR EMBL; AJ011414; CAB56221.1; -.

DR InterPro; IPR001627; -.

DR InterPro; IPR002165; -.

DR Pfam; PF01403; Sema; 2.

DR Pfam; PF01437; Plexin\_repeat; 1.

DR SMART; SM00423; PSI; 1.

KW Signal; Receptor.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 729 PLEXIN-B1/SEP RECEPTOR.

SQ SEQUENCE 729 AA; 78394 MW; 4AE56E0EDBAE6D75 CRC64;

Query Match 60.3%; Score 41; DB 4; Length 729;  
Best Local Similarity 50.0%; Pred. No. 73;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLLCPDRPLE 12  
:|::||| | | |

Db 584 SGVMCPSDPSE 595

RESULT 13  
Q9UJ92  
ID Q9UJ92 PRELIMINARY; PRT; 1952 AA.  
AC Q9UJ92;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PLEXIN-B1/SEP RECEPTOR PRECURSOR.  
GN PLEXIN-B1/SEP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99449305; PubMed=10520995;

RA Tamagnone L., Artigiani S., Chen H., He Z., Ming G.L., Song H.L.,  
Chedotal A., Winberg M.L., Goodman C.S., Poo M.M., Tessier-Lavigne M.,  
Comoglio P.M.;

RT "Plexins are a large family of receptors for transmembrane, secreted  
and GPI-anchored semaphorins in vertebrates.";

RL Cell 99:71-80(1999).

DR EMBL; AJ011415; CAB56222.1; -.

DR InterPro; IPR000566; -.

DR InterPro; IPR001627; -.

DR InterPro; IPR002165; -.

DR InterPro; IPR002569; -.

DR InterPro; IPR002909; -.

DR InterPro; IPR003006; -.

DR Pfam; PF01403; Sema; 2.

DR Pfam; PF01437; Plexin\_repeat; 2.

DR Pfam; PF01833; TIG; 3.

DR PRODOM; PD003489; -; 1.

DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.

DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.

DR SMART; SM00423; PSI; 1.

KW Signal; Receptor.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 1952 PLEXIN-B1/SEP RECEPTOR.

SQ SEQUENCE 1952 AA; 214409 MW; FCAD0630E128EE9C CRC64;

Query Match 60.3%; Score 41; DB 4; Length 1952;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AGLLCPDRPLE 12  
:|::||| | | |

Db 584 SGVMCPSDPSE 595

RESULT 14

O43157

ID O43157 PRELIMINARY; PRT; 2135 AA.

AC O43157;

DT 01-JUN-1998 (TReMBLrel. 06, Created)

DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE KIAA0407.

GN KIAA0407.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RA Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,  
Tanaka A., Kotani H., Nomura N., Ohara O.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB007867; BAA23703.1; -.

DR InterPro; IPR000566; -.

DR InterPro; IPR001627; -.

DR InterPro; IPR001936; -.

DR InterPro; IPR002165; -.

DR InterPro; IPR002909; -.

DR InterPro; IPR003006; -.

DR Pfam; PF01403; Sema; 2.

DR Pfam; PF01437; Plexin\_repeat; 2.

DR Pfam; PF01833; TIG; 3.

DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.

DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.

DR PROSITE; PS0018; RAS\_GTPASE\_ACTIV\_2; 1.

DR SMART; SM00429; IPT; 1.

SQ SEQUENCE 2135 AA; 232295 MW; 12A81B68AF1D340F CRC64;

Query Match 60.3%; Score 41; DB 4; Length 2135;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;



Best Local Similarity 66.7%; Pred. No. 35;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11  
|:|||||:  
Db 169 LVCPDRSI 177

RESULT 18  
O86349 PRELIMINARY; PRT; 261 AA.  
AC O86349;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 27.4 KDA PROTEIN.  
GN RV1457C OR MTV007.04C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA Parkhill J.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL021184; CAA15987.1; -.  
DR TuberculList; RV1457c; -.  
DR InterPro; IPR000005; -.  
DR PROSITE; PS00041; HTH\_ARAC\_FAMILY\_1; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 261 AA; 27370 MW; FF2074B7B862C021 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 261;  
Best Local Similarity 70.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLCPDPRP 10  
||| |||||  
Db 11 AGTFSPDPRP 20

RESULT 19  
Q68387 PRELIMINARY; PRT; 328 AA.  
ID Q68387;  
AC Q68387;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE ORF UL135.  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TOLEDO;

RX MEDLINE=96099416; PubMed=85233595;  
RA Cha T.A., Tom E., Kemble G.W., Duke G.M., Mocarski E.S., Spaete R.R.;  
RT "Human cytomegalovirus clinical isolates carry at least 19 genes not  
RT found in laboratory strains.";  
RL J. Virol. 70:78-83(1996).  
DR EMBL; U33331; AAA85874.1; -.  
SQ SEQUENCE 328 AA; 35735 MW; E5FBF20F5EB6841A CRC64;

Query Match 58.8%; Score 40; DB 14; Length 328;  
Best Local Similarity 77.8%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10  
||| |||||  
Db 273 GLSCPCPRP 281

RESULT 20  
Q9L1A1 PRELIMINARY; PRT; 355 AA.  
ID Q9L1A1;  
AC Q9L1A1;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE PUTATIVE OXIDOREDUCTASE (FRAGMENT).  
GN SC10G8.01C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL158057; CAB76275.1; -.  
FT NON\_TER 355 355  
SQ SEQUENCE 355 AA; 38866 MW; 936B0C4EDB499062 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 355;  
Best Local Similarity 87.5%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPDRP 10  
||| |||||  
Db 235 LLCVDPDRP 242

RESULT 21  
Q9RNX2 PRELIMINARY; PRT; 522 AA.  
ID Q9RNX2;  
AC Q9RNX2;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
DE PUTATIVE EFFLUX PUMP COMPONENT MTRF.  
GN MTRF.



```
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RA Veal W.L., Shafer W.M.; required for high-level antimicrobial
RT resistance mediated by the gonococcal mtr efflux pump."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF176820; AAD51362.2; -
DR InterPro; IPR001589; -
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
SQ SEQUENCE 522 AA; 56131 MW; 1351EC9A32E6AA43 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 522;
Best Local Similarity 70.0%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRPL 11
Db 48 GLSVDPDPV 57

RESULT 22
Q9RNX1 ID Q9RNX1 PRELIMINARY; PRT; 522 AA.
AC Q9RNX1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE PUTATIVE EFFLUX PUMP COMPONENT MTRF.
GN MTRF.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EU75;
RA Veal W.L., Shafer W.M.;
RT "A novel protein (MtrF) required for high-level antimicrobial
RL resistance mediated by the gonococcal mtr efflux pump."
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF176821; AAD51363.2; -
DR InterPro; IPR001589; -
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
SQ SEQUENCE 522 AA; 56175 MW; 7A7BF0BEF128F7E9 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 522;
Best Local Similarity 70.0%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRPL 11
Db 48 GLSVDPDPV 57

RESULT 23
Q9JY63 ID Q9JY63 PRELIMINARY; PRT; 522 AA.
AC Q9JY63;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE EFFLUX PUMP COMPONENT MTRF.
GN NMB1719.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
SQ
```

```
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Oin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
RL EMBL: AE002522; AAF62337.1; -
DR TIGR; NMB1719; -
DR InterPro; IPR001589; -
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
SQ SEQUENCE 522 AA; 56240 MW; 41CCE4AC20E190D1 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 522;
Best Local Similarity 70.0%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRPL 11
Db 48 GLSVDPDPV 57

RESULT 24
Q9JT47 ID Q9JT47 PRELIMINARY; PRT; 522 AA.
AC Q9JT47;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
GN NMA1973.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
RL EMBL: AL162757; CAB85193.1; -
DR InterPro; IPR001589; -
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
SQ SEQUENCE 522 AA; 56139 MW; 412A9DF454BE94DC CRC64;

Query Match 58.8%; Score 40; DB 2; Length 522;
Best Local Similarity 70.0%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRPL 11
Db 48 GLSVDPDPV 57

RESULT 25
O49602 ID O49602 PRELIMINARY; PRT; 528 AA.
AC O49602;
```

DT 01-JUN-1998 (TReMBLrel. 06, Created)  
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE IMPORTIN ALPHA-LIKE PROTEIN (FRAGMENT).  
GN IMPA-4.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=VEGETATIVE TISSUE;  
RA Schledz M., Leclerc D., Neuhaus G., Merkle T.;  
RL Plant Physiol. 116:868-868(1998).  
DR EMBL; Y14616; CAA74966.1; -.  
DR HSSP; Q02821; 1BK5.  
DR Mendel; 24707; Arath;2729;24707.  
DR InterPro; IPR000225; -.  
DR InterPro; IPR002652; -.  
DR InterPro; IPR003006; -.  
DR Pfam; PF00514; Armadillo\_seg; 8.  
DR Pfam; PF01749; IBB; 1.  
DR PROSITE; PS50176; ARM\_REPEAT; 2.  
DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.  
DR SMART; SM00185; ARM; 1.  
FT NON\_TER 1  
SQ SEQUENCE 528 AA; 58265 MW; D794CA1CE959E731 CRC64;  
  
Query Match 58.8%; Score 40; DB 10; Length 528;  
Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 LLCPPDR 9  
|:|||||  
Db 418 LICPPDR 424  
  
RESULT 26  
O80480  
ID O80480 PRELIMINARY; PRT; 538 AA.  
AC O80480;  
DT 01-NOV-1998 (TReMBLrel. 08, Created)  
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE T12M4.2 PROTEIN.  
DE T12M4.2.  
GN T12M4.2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Oji, O, Kwan, A,  
RA Liu S., Li J., Araujo R., Au M., Brendel V., Buehler E., Conway A.B.,  
RA Conway A.R., Dewar K., Feng J., Kim C., Kurtz D., Li Y., Palm C.J.,  
RA Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,  
RA Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC T12M4 sequence, complete  
sequence."  
RT Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AC003114; AAC24079.1; -.  
DR HSSP; Q02821; 1BK5.  
DR Mendel; 31141; Arath;2729;31141.  
DR InterPro; IPR000225; -.  
DR InterPro; IPR002652; -.  
DR Pfam; PF00514; Armadillo\_seg; 8.  
DR Pfam; PF01749; IBB; 1.  
DR PROSITE; PS50176; ARM\_REPEAT; 3.  
DR SMART; SM00185; ARM; 1.  
SQ SEQUENCE 538 AA; 59445 MW; C342198841B87CCA CRC64;  
  
Query Match 58.8%; Score 40; DB 10; Length 538;  
Best Local Similarity 85.7%; Pred. No. 81;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 LLCPPDR 9  
|:|||||  
Db 428 LICPPDR 434  
  
RESULT 27  
Q9RRT1  
ID Q9RRT1 PRELIMINARY; PRT; 633 AA.  
AC Q9RRT1;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE ABC TRANSPORTER, ATP-BINDING PROTEIN, MSBA FAMILY.  
DE DR2404.  
GN Deinococcus radiodurans.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1."  
RL Science 286:1571-1577(1999).  
DR EMBL; AE002071; AAF11948.1; -.  
DR HSSP; P13569; 1NBD.  
DR TIGR; DR2404; -.  
DR InterPro; IPR001687; -.  
DR InterPro; IPR003439; -.  
DR InterPro; IPR003593; -.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR SMART; SM00382; AAA; 1.  
KW ATP-binding.  
SQ SEQUENCE 633 AA; 69337 MW; DB8549554C15EB32 CRC64;  
  
Query Match 58.8%; Score 40; DB 2; Length 633;  
Best Local Similarity 87.5%; Pred. No. 94;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 LCPDPRPL 11  
|:|||||  
Db 369 LAPDPRPL 376



RA Theologis A.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC000104; AAB70424.1; -;  
DR Mendel; 12684; Arath;1884;12684.  
KW Hypothetical protein.  
SQ SEQUENCE 399 AA; 44689 MW; 680ED6D4BA3ED347 CRC64;

Query Match 58.1%; Score 39.5; DB 10; Length 399;  
Best Local Similarity 63.6%; Pred. No. 75;  
Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 2 GLLCPDP-RPL 11  
|| |||| :|:  
Db 252 GLFCPDPPIKPI 262

RESULT 31  
Q63552 PRELIMINARY; PRT; 146 AA.  
AC Q63552;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE SMR1-ALPHA1.  
GN VCS-ALPHA1.

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=101116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SUBMANDIBULAR GLAND;  
RX MEDLINE=95169272; PubMed=7865131;  
RA Singer M., Courty Y., Rougeon F.;  
RT "Recent evolution of genes encoding the prohormone-like protein SMR1  
in the rat submandibular gland.";  
RL DNA Cell Biol. 14:137-144(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SUBMANDIBULAR GLAND;  
RX Courtney Y., Rosinski-Chupin I., Rougeon F.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.  
RC TISSUE=SUBMANDIBULAR GLAND;  
RX MEDLINE=96276306; PubMed=8754212;  
RA Courtney Y., Singer M., Rosinski-Chupin I., Rougeon F.;  
RT "Episodic evolution and rapid divergence of members of the rat  
multigene family encoding the salivary prohormone-like protein SMR1.";  
RL Mol. Biol. Evol. 13:758-766(1996).  
DR EMBL; X77819; CAA54834.1; -;  
FT CHAIN 23 146 SMR1-ALPHA1.  
SQ SEQUENCE 146 AA; 16282 MW; BCBC5E360A1C50E8 CRC64;

Query Match 57.4%; Score 39; DB 11; Length 146;  
Best Local Similarity 77.8%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11  
| | | | |  
Db 98 LTAPDRPL 106

RESULT 32  
Q9V3K5 PRELIMINARY; PRT; 262 AA.  
AC Q9V3K5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE BG:DS06874.1 PROTEIN.

GN OS  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=107311132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Beasley E.M.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Bolshakov S.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Brottier P.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, AND CN BW SP;  
RX MEDLINE=99403001; PubMed=10471707;  
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,  
Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,  
Celniker S., Rubin G.M.;  
RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
Drosophila melanogaster: the Adh region.";  
RL Genetics 153:179-219(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y, AND CN BW SP;  
RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,  
Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,  
Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,  
Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,  
Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,  
Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,  
Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,  
Zieran L.L., Rubin G.M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003645; AAF53408.1; -;

DR EMBL; AE003645; AAF53408.1; -;



DR EMBL; AE003411; AAF44891.1; -  
DR FlyBase; FBgn0028870; BG:DS06874.1.  
SQ SEQUENCE 262 AA; 30846 MW; AD2BEDABB81A051C CRC64;

Query Match 57.4%; Score 39; DB 5; Length 262;  
Best Local Similarity 85.7%; Pred. No. 62;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 PDPRLP 12  
Db 235 PDPRLP 241

RESULT 33  
ID Q9JHJ7 PRELIMINARY; PRT; 305 AA.

AC Q9JHJ7;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE TRYPTASE 4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=TESTIS;  
RA Wong G.W.; Stevens R.L.;  
RT "Cloning of the mouse tryptase 4.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=TESTIS;  
RA Wong G.W.; Stevens R.L.;  
RT "Cloning of the mouse tryptase 4.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RA Wong G.W.; Li L.; Stevens R.L.;  
RT "Mouse tryptase 4 gene.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
DR EMBL; AF176209; AAF64407.2; -  
DR EMBL; AF226710; AAF64428.2; -  
DR InterPro; IPR001254; -  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
SQ SEQUENCE 305 AA; 34326 MW; 5B5D0EFB93EFB21C CRC64;

Query Match 57.4%; Score 39; DB 11; Length 305;  
Best Local Similarity 80.0%; Pred. No. 71;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GLLCPDPRPL 11  
Db 280 GLLRPDPVPL 289

RESULT 34  
ID O60393 PRELIMINARY; PRT; 392 AA.

AC O60393;  
DT 01-AUG-1998 (Tremblrel. 07, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE WUGSC:H\_DJ0545C24.1 PROTEIN (FRAGMENT).

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dauphin S.; Biewald T.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004534; AAC12957.1; -  
DR InterPro; IPR001356; -  
DR Pfam; PF00046; homeobox; 2.  
DR PROSITE; PS50071; HOMEBOX\_2; 1.  
DR SMART; SM00389; HOX; 1.  
FT NON\_TER  
SQ SEQUENCE 392 AA; 41949 MW; 435C3061FC4AE474 CRC64;

Query Match 57.4%; Score 39; DB 4; Length 392;  
Best Local Similarity 77.8%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGLLCPPDR 9  
Db 57 AGLLCPPPR 65

RESULT 35  
ID P70225 PRELIMINARY; PRT; 432 AA.

AC P70225; O09074;  
DT 01-FEB-1997 (Tremblrel. 02, Created)  
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE INTERLEUKIN-11 RECEPTOR ALPHA CHAIN 2 PRECURSOR (IL11RA2) (IL-11RBETA)  
DE (INTERLEUKIN-11 RECEPTOR BETA CHAIN).  
DE IL11RA2 OR IL-11RBETA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD1; TISSUE=TESTIS;  
RA Billinski P.; Hall M.A.; Neuhaus H.; Gissel C.; Heath J.K.; Gossler A.;  
RT "Two differentially expressed interleukin-11 receptor genes in the mouse genome.";  
RL Biochem. J. 320:359-363(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1; TISSUE=TESTIS;  
RA Robb L.; Hilton D.J.; Willson T.A.; Begley C.G.;  
RT "Structural analysis of the gene encoding the murine interleukin-11 receptor alpha-chain and a related locus.";  
RL J. Biol. Chem. 271:13754-13761(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1; TISSUE=TESTIS;  
RA Robb L.; Hilton D.J.; Brook-Carter P.T.; Begley C.G.;  
RT "Identification of a second murine interleukin-11 receptor alpha-chain gene (il11ra2) with a restricted pattern of expression.";  
RL Genomics 40:387-394(1997).

CC -1- FUNCTION: PART OF THE RECEPTOR FOR INTERLEUKIN 11.  
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.  
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX

```
CC      DOMAIN.
DR      EMBL; X94157; CAA63872.1; -.
DR      EMBL; X94158; CAA63872.1; JOINED.
DR      EMBL; X94159; CAA63872.1; JOINED.
DR      EMBL; X94160; CAA63872.1; JOINED.
DR      EMBL; X94161; CAA63872.1; JOINED.
DR      EMBL; X98519; CAA67144.1; -.
DR      EMBL; U69491; AAC53114.1; -.
DR      MGD; MGI:109123; I111ra2.
DR      InterPro; IPR001777; -.
DR      InterPro; IPR002996; -.
DR      InterPro; IPR003006; -.
DR      InterPro; IPR003530; -.
DR      Pfam; PF00041; fn3; 2.
DR      Pfam; PF00047; ig; 1.
DR      PROSITE; PS01354; HEMATOPO_REC_L_F3; UNKNOWN_1.
DR      SMART; SM00060; FN3; 1.
KW      Receptor; Transmembrane; Glycoprotein; Immunoglobulin domain; Signal.
FT      SIGNAL      1      23      POTENTIAL.
FT      CHAIN      24      432      INTERLEUKIN-11 RECEPTOR ALPHA CHAIN 2.
FT      DOMAIN      24      367      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM      368      393      POTENTIAL.
FT      DOMAIN      394      432      CYTOPLASMIC (POTENTIAL).
FT      DOMAIN      41      102      IG-LIKE C2-TYPE DOMAIN.
FT      CARBOHYD      127      127      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD      194      194      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CONFLICT      200      200      S -> P (IN CAA63872).
FT      CONFLICT      384      384      V -> L (IN CAA63872).
SQ      SEQUENCE      432 AA; 46721 MW; 0D8E49723EC99EC5 CRC64;

Query Match      57.4%; Score 39; DB 11; Length 432;
Best Local Similarity      77.8%; Pred. No. 97;
Matches      7; Conservative      1; Mismatches      1; Indels      0; Gaps      0;

QY      4 LCPDPRPLE 12
      | |||||:
Db      352 LQDPRPLD 360

RESULT      36
Q64385
ID      Q64385      PRELIMINARY;      PRT;      432 AA.
AC      Q64385;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE      INTERLEUKIN-11 RECEPTOR ALPHA CHAIN 1 PRECURSOR (NR1) (ETL2)
DE      (IL-11RALPHA) (I111RAL).
GN      IL11RAL OR IL11RA OR ETL2 OR ETL2/IL11 REC.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6 X CBA; TISSUE=LIVER;
RX      MEDLINE=95045367; PubMed=7957045;
RA      Hilton D.J., Hilton A.A., Raicevic A., Rakar S., Harrison-Smith M.,
RA      Gough N.M., Begley C.G., Metcalf D., Nicola N.A., Willson T.A.;
RT      "Cloning of a murine IL-11 receptor alpha-chain; requirement for gp130
RT      for high affinity binding and signal transduction.";
RL      EMBO J. 13:4765-4775(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BALB/C, AND C57BL/6; TISSUE=EMBRYO;
RA      Neuhaus H., Bettenhausen B., Bilinski P., Simon-Chazottes D.,
RA      Guenet J.L., Gossler A.;
RL      Dev. Biol. 166:521-542(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BALB/C, AND C57BL/6;
RA      Gossler A.;
```

```
RL      Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN      [4]
RX      MEDLINE=97129000; PubMed=8973540;
RA      Bilinski P., Hall M.A., Neuhaus H., Gissel C., Heath J.K., Gossler A.;
RT      "Two differentially expressed interleukin-11 receptor genes in the
RT      mouse genome.";
RL      Biochem. J. 320:359-363(1996).
CC      -!- FUNCTION: PART OF THE RECEPTOR FOR INTERLEUKIN 11. BINDS TO IL-11
CC      WITH LOW AFFINITY, BUT DOES NOT TRANSDUCE A SIGNAL.
CC      -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC      -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC      -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC      ONE IG-LIKE DOMAIN.
CC      -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
DR      EMBL; X74953; CAA52908.1; -.
DR      EMBL; U14412; AAA53248.1; -.
DR      EMBL; X94162; CAA63873.1; -.
DR      EMBL; X94163; CAA63873.1; JOINED.
DR      MGD; MGI:107426; I111ral.
DR      InterPro; IPR001777; -.
DR      InterPro; IPR002996; -.
DR      InterPro; IPR003006; -.
DR      InterPro; IPR003530; -.
DR      Pfam; PF00041; fn3; 2.
DR      Pfam; PF00047; ig; 1.
DR      PROSITE; PS01354; HEMATOPO_REC_L_F3; UNKNOWN_1.
DR      SMART; SM00060; FN3; 1.
KW      Receptor; Transmembrane; Glycoprotein; Immunoglobulin domain; Signal.
FT      SIGNAL      1      23      POTENTIAL.
FT      CHAIN      24      432      INTERLEUKIN-11 RECEPTOR ALPHA CHAIN 1.
FT      DOMAIN      24      367      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM      368      393      POTENTIAL.
FT      DOMAIN      394      432      CYTOPLASMIC (POTENTIAL).
FT      DOMAIN      41      102      IG-LIKE C2-TYPE DOMAIN.
FT      CARBOHYD      127      127      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD      194      194      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE      432 AA; 46655 MW; 068389943502BBFC CRC64;

Query Match      57.4%; Score 39; DB 11; Length 432;
Best Local Similarity      77.8%; Pred. No. 97;
Matches      7; Conservative      1; Mismatches      1; Indels      0; Gaps      0;

QY      4 LCPDPRPLE 12
      | |||||:
Db      352 LQDPRPLD 360

RESULT      37
O49599
ID      O49599      PRELIMINARY;      PRT;      532 AA.
AC      O49599;
DT      01-JUN-1998 (TREMBLrel. 06, Created)
DT      01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT      01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE      IMPORTIN ALPHA-LIKE PROTEIN.
GN      IMPAL.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC      Brassicales; Brassicaceae; Arabidopsiis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=VEGETATIVE;
RA      Schledz M., Leclerc D., Neuhaus G., Merkle T.;
RL      Plant Physiol. 116:868-868(1998).
DR      EMBL; Y15224; CAA75513.1; -.
DR      HSSP; Q02821; 1BK5.
DR      Mendel; 24587; Arath; 2729; 24587.
DR      InterPro; IPR000225; -.
```

DR InterPro: IPR002652; -  
DR Pfam: PF00514; Armadillo\_seg; 8.  
DR Pfam: PF01749; IBB; 1. REPEAT; 4.  
DR PROSITE: PS50176; ARM\_REPEAT; 1.  
DR SMART: SM00185; ARM; 1.  
SQ SEQUENCE 532 AA; 58670 MW; DA80EC17E1ACF70E CRC64;

Query Match 57.4%; Score 39; DB 10; Length 532;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9  
Db 421 LVCPDPR 427

RESULT 38  
ID O81520 PRELIMINARY; PRT; 532 AA.  
AC O81520;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE IMPORTIN ALPHA.  
GN F3E22.14.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=96373199; PubMed=8776900;  
RA Hicks G.R.; Smith H.M.; Lobreau S.; Raikhel N.V.;  
RT "Nuclear import in permeabilized protoplasts from higher plants has  
RT unique features.";  
RL Plant Cell 8:1337-1352(1996).

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=97336305; PubMed=9193081;  
RA Smith H.M.; Hicks G.R.; Raikhel N.V.;  
RT "Importin alpha from Arabidopsis thaliana is a nuclear import receptor  
RT that recognizes three classes of import signals.";  
RL Plant Physiol. 114:411-417(1997).

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Raikhel N.V.; Smith H.M.S.;  
RN Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Lin X.; Kaul S.; Town C.D.; Benito M.; Creasy T.H.; Haas B.; Wu D.;  
RA Maiti R.; Ronning C.M.; Koo H.; Fujii C.Y.; Utterback T.R.;  
RA Barnstead M.E.; Bowman C.L.; White O.; Nierman W.C.; Fraser C.M.;  
RT "Arabidopsis thaliana chromosome III BAC F3E22 genomic sequence.";  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF077528; AAC27644.1; -  
DR EMBL: AC023912; AAF63826.1; -  
DR HSSP: Q02821; 1BK5.  
DR Mendel; 32157; Arath; 2729; 32157.  
DR InterPro: IPR00225; -  
DR InterPro: IPR002652; -  
DR Pfam: PF00514; Armadillo\_seg; 8.  
DR Pfam: PF01749; IBB; 1.  
DR PROSITE: PS50176; ARM\_REPEAT; 4.  
DR SMART: SM00185; ARM; 1.  
SQ SEQUENCE 532 AA; 58644 MW; 9425EACB98D585C7 CRC64;

Query Match 57.4%; Score 39; DB 10; Length 532;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9  
Db 421 LVCPDPR 427

RESULT 39  
ID Q9SLX0 PRELIMINARY; PRT; 534 AA.  
AC Q9SLX0;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE IMPORTIN ALPHA IB.  
GN Oryza sativa (Rice).

OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;  
OC Oryza.  
OX NCBI\_TaxID=4530;

RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE; TISSUE=LEAF;  
RA Jiang C.; Shoji K.; Matsuki R.; Inagaki N.; Ban H.; Iwasaki T.;  
RA Imamoto N.; Yoneda Y.; Yamamoto N.;  
RT "Cloning and functional analysis of a novel improtin alpha homologue  
RT from rice (Oryza sativa).";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB024311; BAA88950.1; -  
DR HSSP: Q02821; 1BK5  
DR InterPro: IPR000225; -  
DR InterPro: IPR002652; -  
DR Pfam: PF00514; Armadillo\_seg; 8.  
DR Pfam: PF01749; IBB; 1.  
DR PROSITE: PS50176; ARM\_REPEAT; 3.  
DR SMART: SM00185; ARM; 1.  
SQ SEQUENCE 534 AA; 58532 MW; 90C8A3F7ADEF1720 CRC64;

Query Match 57.4%; Score 39; DB 10; Length 534;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9  
Db 427 LVCPDPR 433

RESULT 40  
ID O49600 PRELIMINARY; PRT; 535 AA.  
AC O49600;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE IMPORTIN ALPHA-LIKE PROTEIN.  
GN IMPA-2.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RP SEQUENCE FROM N.A.  
RC TISSUE=VEGETATIVE TISSUE;  
RA Schledz M.; Leclerc D.; Neuhaus G.; Merkle T.;  
RL Plant Physiol. 116:868-868(1998).  
DR EMBL: Y14615; CAA74965.1; -  
DR HSSP: Q02821; 1BK5.  
DR Mendel; 24588; Arath; 2729; 24588.  
DR InterPro: IPR000225; -

DR InterPro; IPR002652; -  
DR Pfam; PF00514; Armadillo\_seg; 8.  
DR Pfam; PF01749; IBB; 1.  
DR PROSITE; PS50176; ARM\_REPEAT; 3.  
DR SMART; SM00185; ARM; 1.  
SQ SEQUENCE 535 AA; 58858 MW; 029745BF1DB5A5A9 CRC64;

Query Match 57.4%; Score 39; DB 10; Length 535;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9  
I:|||||

Db 426 LVCPDR 432

RESULT 41  
Q9P200 PRELIMINARY; PRT; 740 AA.  
AC Q9P200;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE KIAA1528 PROTEIN (FRAGMENT).  
GN KIAA1528.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20277482; PubMed=10819331;  
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human  
genes.XVII.The complete sequences of 100 new cDNA clones from brain  
which code for large proteins in vitro.";  
RL DNA Res. 7:143-150(2000).

DR EMBL; AB040961; BAA96052.1; -.  
DR InterPro; IPR001304; -.  
DR InterPro; IPR001841; -.  
DR Pfam; PF00097; zf-C3HC4; 1.  
DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 740 AA; 79828 MW; 9D8B622DFF99397D CRC64;

Query Match 57.4%; Score 39; DB 4; Length 740;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CPDPRPL 11  
I:|||||

Db 22 CPQPRPL 28

RESULT 42  
P91658 PRELIMINARY; PRT; 974 AA.  
AC P91658;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE FURROWED.  
GN FW OR CG1500.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Leshko-Lindsay L., Corces V.G.;

RL Development 0:0-0(0).  
DR EMBL; U70770; AAB36703.1; -.  
DR HSSP; P02749; IQUB.  
DR FlyBase; FBgn0001083; fw.  
DR InterPro; IPR000436; -.  
DR InterPro; IPR001304; -.  
DR Pfam; PF00059; lectin\_c; 1.  
DR Pfam; PF00084; sushi; 10.  
DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; UNKNOWN\_1.  
DR PROSITE; PS50041; C\_TYPE\_LECTIN\_2; 1.  
DR SMART; SM00032; CCP; 1.  
SQ SEQUENCE 974 AA; 107557 MW; A406335FB6726757 CRC64;

Query Match 57.4%; Score 39; DB 5; Length 974;  
Best Local Similarity 71.4%; Pred. No. 2e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 CPDPRPL 11  
I:|||||

Db 293 CPDPQPI 299

RESULT 43  
Q9VYR4 PRELIMINARY; PRT; 1124 AA.  
ID Q9VYR4;  
AC Q9VYR4;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE FW GENE PRODUCT.  
GN FW OR CG1500.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpis G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,



RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Zheng L.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003487; AAF48125.1; -  
DR HSSP; P02749; 1QUB.  
DR FlyBase; FBgn001083; fw.  
DR InterPro; IPR000436; -  
DR InterPro; IPR001304; -  
DR Pfam; PF00059; lectin\_c; 1.  
DR Pfam; PF00084; sushi; 11.  
DR PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
DR PROSITE; PS50041; C-TYPE\_LECTIN\_2; 1.  
DR SMART; SM00032; CCP; 1.  
SQ SEQUENCE 1124 AA; 123051 MW; C0EB0528A415F34A CRC64;

Query Match 57.4%; Score 39; DB 5; Length 1124;  
Best Local Similarity 71.4%; Pred. No. 2.3e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 CPDPRPL 11  
Db 475 CPDPQPI 481

RESULT 44  
QY7C6 PRELIMINARY; PRT; 1529 AA.  
AC QY7C6;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE ESTERASE.  
OS Aspergillus terreus.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=33178;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC20542;  
RA Kennedy J., Auclair K., Kendrew S.G., Park C., Vederas J.C.,  
RA Hutchinson C.R.;  
RT "Accessory Proteins Modulate Polyketide Synthase Activity During  
RT Lovastatin Biosynthesis.";  
RL Science 0:0-0(1999).  
DR EMBL; AF141924; AAD34550.1; -  
DR InterPro; IPR001087; -  
SQ SEQUENCE 1529 AA; 170505 MW; 456FB780A6B531A3 CRC64;

Query Match 57.4%; Score 39; DB 3; Length 1529;  
Best Local Similarity 77.8%; Pred. No. 3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPRPL 11  
Db 1097 LLCPPNLL 1105

RESULT 45  
Q9JIO9 PRELIMINARY; PRT; 1734 AA.  
AC Q9JIO9;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE RIM BINDING PROTEIN 1B (FRAGMENT).  
GN RBPIB.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20347919; PubMed=10748113;  
RA Wang Y., Sugita S., Sudhof T.C.;  
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with  
RT Rab3 and a new class of Src homology 3 domain proteins.";  
RL J. Biol. Chem. 275:20033-20044(2000).  
DR EMBL; AF199338; AAF81660.1; -  
DR InterPro; IPR001452; -  
DR InterPro; IPR001777; -  
DR Pfam; PF00018; SH3; 3.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PROSITE; PS50002; SH3; 3.  
DR SMART; SM00060; FN3; 1.  
DR NON\_TER 1  
SQ SEQUENCE 1734 AA; 188524 MW; B771ECBE900A0AC7 CRC64;

Query Match 57.4%; Score 39; DB 11; Length 1734;  
Best Local Similarity 75.0%; Pred. No. 3.4e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11  
Db 139 LCPDPPV 146

Search completed: June 28, 2001, 11:55:32  
Job time: 153 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 66.55 Seconds  
(without alignments)  
14.575 Million cell updates/sec

Title: US-09-439-313-554  
Perfect score: 82  
Sequence: 1 YVPPLLLLEVGVEKFM 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	553	19 AAW71869	Amino acid encoded
2	82	100.0	553	19 AAW69385	Prostate tumour sp
3	82	100.0	553	21 AAB28527	Protein encoded by
4	82	100.0	553	21 AAY82002	Human immunogenic
5	43	52.4	345	21 AAG11269	Arabidopsis thalia
6	43	52.4	345	21 AAG51470	Arabidopsis thalia
7	43	52.4	351	21 AAG11268	Arabidopsis thalia
8	43	52.4	351	21 AAG51469	Arabidopsis thalia
9	43	52.4	376	21 AAG11267	Arabidopsis thalia
10	43	52.4	376	21 AAG51468	Arabidopsis thalia
11	41	50.0	3472	21 AAY90913	Cenarchaeum symbio

12	40	48.8	63	18	AAV11325
13	40	48.8	163	21	AAV70541
14	40	48.8	172	19	AAV86095
15	40	48.8	437	16	AAR82839
16	40	48.8	695	20	AAV20069
17	40	48.8	718	20	AAV20068
18	39	47.6	101	21	AAV79062
19	39	47.6	643	21	AAV87121
20	39	47.6	674	21	AAV87209
21	38	46.3	58	21	AAB44374
22	38	46.3	119	21	AAB42741
23	38	46.3	298	22	AAB50540
24	38	46.3	312	22	AAB80121
25	37	45.1	119	11	AAR09304
26	37	45.1	398	13	AAR20504
27	37	45.1	398	13	AAR20505
28	37	45.1	398	13	AAR20506
29	37	45.1	398	13	AAR20507
30	37	45.1	401	16	AAR83104
31	37	45.1	401	16	AAR83105
32	37	45.1	401	16	AAR83106
33	37	45.1	402	14	AAR44435
34	37	45.1	402	14	AAR42737
35	37	45.1	402	14	AAR42738
36	37	45.1	402	14	AAR42739
37	37	45.1	402	14	AAR42740
38	37	45.1	402	16	AAR83101
39	37	45.1	402	16	AAR83102
40	37	45.1	402	16	AAR83103
41	37	45.1	402	16	AAR67259
42	37	45.1	476	13	AAR20503
43	37	45.1	476	14	AAR42736
44	37	45.1	476	16	AAR82250
45	37	45.1	1410	21	AAG31539
46	37	45.1	1455	21	AAG31538
47	37	45.1	1465	21	AAG31537
48	37	45.1	1693	12	AAR14618
49	36	43.9	83	21	AAG41140
50	36	43.9	90	21	AAG41139
51	36	43.9	209	22	AAB60348
52	36	43.9	307	20	AAV35623
53	36	43.9	353	21	AAG32571
54	36	43.9	353	21	AAG49006
55	36	43.9	354	21	AAG09010
56	36	43.9	371	21	AAG32570
57	36	43.9	371	21	AAG49005
58	36	43.9	372	21	AAG09009
59	36	43.9	458	21	AAG32569
60	36	43.9	458	21	AAG49004
61	36	43.9	459	21	AAG09008
62	36	43.9	516	20	AAV55928
63	36	43.9	520	20	AAV01496
64	36	43.9	520	20	AAV01497
65	36	43.9	625	18	AAW09929

ALIGNMENTS

RESULT 1	
AAW71869	
ID	AAW71869 standard; Protein; 553 AA.
XX	
AC	AAW71869;
XX	
DT	06-JAN-1999 (first entry)
XX	
DE	Amino acid encoded by prostate tumour clone L1-12.
XX	
KW	Prostate; cancer; tumour; vaccine; immunogen; clone.
XX	
OS	Homo sapiens.
XX	

PN WO9837093-A2.  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03492.  
XX  
PR 09-FEB-1998; 98US-0020956.  
PR 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
DR WPI; 1998-609886/51.  
DR N-PSDB; AAV61201.  
XX  
PT Polypeptides comprising immunogenic portions of prostate proteins  
PT used in a vaccine for the treatment of prostate cancer  
XX  
PS Example 1; Page 82-84; 130pp; English.  
XX  
CC The present sequence is an immunogenic portion of a prostate tumour  
CC protein. The immunogen, or the DNA encoding it, can be used as a  
CC vaccine for the treatment of prostate cancer. The immunogen was  
CC isolated from a prostate tumour cDNA library obtained by subtracting  
CC a prostate tumour cDNA expression library with a normal tissue cDNA  
CC library.  
XX  
SQ Sequence 553 AA;  
  
Query Match 100.0%; Score 82; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. NO. 3.7e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YVPPLLLLEVGVEEKFM 16  
Db | | | | | | | | | | | | | | | |  
38 yvppllllevgveekfm 53  
  
RESULT 2  
AAW69385  
ID AAW69385 standard; Protein; 553 AA.  
XX  
AC AAW69385;  
XX  
DT 08-DEC-1998 (first entry)  
XX  
DE Prostate tumour specific gene clone L1-12 protein.  
XX  
DE Prostate tumour specific gene; human; prostate cancer; detection;  
KW therapy.  
KW Homo sapiens.  
XX  
OS WO9837418-A2.  
XX  
PN 27-AUG-1998.  
XX  
PD 25-FEB-1998; 98WO-US03690.  
XX  
PF 09-FEB-1998; 98US-0904809.  
PR 25-FEB-1997; 97US-0806596.  
PR 01-AUG-1997; 97US-0904809.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Dillon DC, Xu J;  
PI  
XX WPI; 1998-480805/41.  
DR N-PSDB; AAV58586.  
XX

PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
XX  
PS Example 1; Page 87-89; 141pp; English.  
XX  
CC This sequence is encoded by a human prostate tumour specific gene, and  
CC can be used in the method of the invention. The method is for detecting  
CC prostate cancer comprising contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC this protein sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect;  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
XX  
SQ Sequence 553 AA;  
  
Query Match 100.0%; Score 82; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. NO. 3.7e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YVPPLLLLEVGVEEKFM 16  
Db | | | | | | | | | | | | | | | |  
38 yvppllllevgveekfm 53  
  
RESULT 3  
AAB28527  
ID AAB28527 standard; Protein; 553 AA.  
XX  
AC AAB28527;  
XX  
DT 07-FEB-2001 (first entry)  
XX  
DE Protein encoded by human breast tumour cDNA clone P501S.  
XX  
DE Human; breast tumour antigen; cytostatic; immunotherapy;  
KW breast cancer; vaccine.  
KW Homo sapiens.  
XX  
OS WO2000061756-A2.  
XX  
PN 19-OCT-2000.  
XX  
PD 10-APR-2000; 2000WO-US09688.  
XX  
PF 09-APR-1999; 99US-0288950.  
PR 02-JUL-1999; 99US-0346327.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Reed SG, Xu J, Dillon DC;  
PI  
XX WPI; 2000-638568/61.  
DR N-PSDB; AAC79473.  
XX  
PT A novel isolated polypeptide comprising an immunogenic portion of a  
PT breast cancer protein useful in the detection and treatment of breast  
PT cancer -  
XX  
PS Claim 2; Page 92-93; 95pp; English.  
XX  
CC The present sequence is encoded by a cDNA sequence which was isolated  
CC from a breast tumour cDNA library. It is provided in a specification  
CC relating to compounds for immunotherapy and diagnosis of breast cancer.  
CC Breast tumour antigens and the polynucleotides that encode them may be  
CC used in the production of a pharmaceutical composition to be used in the  
CC treatment of breast cancer. Proliferated T cells and incubated antigen  
CC presenting cells are also required. The polypeptides and polynucleotides  
CC may also be used to produce a vaccine.  
XX



SQ Sequence 553 AA;

Query Match 100.0%; Score 82; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 3.7e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YVPPLLEVGVEKFM 16  
Db 38 yvpplllevgveekfm 53

RESULT 4  
AAY82002 ID AAY82002 standard; Protein; 553 AA.  
XX AC AAY82002;  
DT 13-JUN-2000 (first entry)  
DE Human immunogenic prostate tumour protein sequence SEQ ID NO:113.  
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
KW immunogenic; cytostatic; vaccine.  
OS Homo sapiens.  
PN WO200004149-A2.  
PD 27-JAN-2000.  
PF 14-JUL-1999; 99WO-US15838.  
PR 14-JUL-1998; 98US-O115453.  
PR 14-JUL-1998; 98US-O116134.  
PR 23-SEP-1998; 98US-O159812.  
PR 23-SEP-1998; 98US-O159822.  
PR 15-JAN-1999; 99US-O232149.  
PR 15-JAN-1999; 99US-O232880.  
PR 09-APR-1999; 99US-O288946.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;  
DR WPI; 2000-171268/15.  
PT New polypeptide useful for treating and diagnosing prostate cancer  
PT comprises an immunogenic portion of prostate tumor protein -  
PS Claim 3; Page 138-139; 263pp; English.

The present invention describes isolated polypeptides, comprising an immunogenic portion of a prostate tumour protein (PTP). The polypeptides and polynucleotides encoding them have cytotstatic activity and can be used in vaccines and in gene therapy. The polypeptides and polynucleotides encoding them, antigen presenting cells which express the polypeptides, antibodies against the polypeptides and vaccines comprising them can be used for inhibiting the development of prostate cancer in a patient. The polypeptides can be used to generate antibodies or anti-idiotypic antibodies for passive immuno therapy. A portion of the polynucleotides encoding the polypeptides can be used as a probe or to modulate the expression of the polypeptides. AAA06241 to AAA06691 and AAY82000 to AAY82020 represent sequences used in the exemplification of the present invention.

SQ Sequence 553 AA;

PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 52.4%; Score 43; DB 21; Length 345;  
Best Local Similarity 61.5%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 3 PPLLEVGVEEKF 15  
Db 118 ppilqngqeenf 130  
RESULT 6  
AAG51470

ID AAG51470 standard; Protein; 345 AA.  
XX  
AC AAG51470;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65328.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.





PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 52.4%; Score 43; DB 21; Length 351;

Best Local Similarity 61.5%; Pred. No. 14;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PPLLLEGVVEEKF 15  
|||::| || |  
Db 124 pplilngqeenf 136

RESULT 8

AAG51469

ID AAG51469 standard; Protein; 351 AA.

XX

AC AAG51469;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 65327.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX

OS Arabidopsis thaliana.

XX



PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	
PR	28-SEP-1999;	99US-0156458.	
PR	29-SEP-1999;	99US-0156596.	
PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Query Match 52.4%; Score 43; DB 21; Length 351;			
Best Local Similarity 61.5%; Pred. No. 14;			
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;			
QY	3	PPLLLEVGVEEKF 15	
		:	
Db	124	pplilqnggeenf 136	
RESULT 9			
AAG11267			
ID	AAG11267 standard; Protein; 376 AA.		
XX			
AC	AAG11267;		
XX			
DT	17-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 9920.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999; 99US-0121825.		

PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.

PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.	Query Match 52.4%; Score 43; DB 21; Length 376;		
PR	05-AUG-1999;	99US-0147260.	Best Local Similarity 61.5%; Pred. NO. 15;		
PR	06-AUG-1999;	99US-0147303.	Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;		
PR	06-AUG-1999;	99US-0147416.	OY 3 PPLLEVGVEEFK 15		
PR	09-AUG-1999;	99US-0147493.	Db 149 pplilqngqeenf 161		
PR	09-AUG-1999;	99US-0147935.	RESULT 10		
PR	10-AUG-1999;	99US-0148171.	AAG51468		
PR	11-AUG-1999;	99US-0148319.	ID AAG51468 standard; Protein; 376 AA.		
PR	12-AUG-1999;	99US-0148341.	XX AC AAG51468;		
PR	13-AUG-1999;	99US-0148565.	XX DT 18-OCT-2000 (first entry)		
PR	13-AUG-1999;	99US-0148565.	XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 65326.		
PR	16-AUG-1999;	99US-0148684.	XX KW Protein identification; signal transduction pathway; metabolic pathway;		
PR	17-AUG-1999;	99US-0149368.	KW hybridisation assay; genetic mapping; gene expression control; promoter;		
PR	17-AUG-1999;	99US-0149175.	XX KW termination sequence.		
PR	18-AUG-1999;	99US-0149426.	XX OS Arabidopsis thaliana.		
PR	20-AUG-1999;	99US-0149722.	XX XX EP1033405-A2.		
PR	20-AUG-1999;	99US-0149723.	XX PD 06-SEP-2000.		
PR	20-AUG-1999;	99US-0149929.	XX PF 25-FEB-2000; 2000EP-0301439.		
PR	23-AUG-1999;	99US-0149902.	XX XX 25-FEB-1999; 99US-0121825.		
PR	23-AUG-1999;	99US-0149930.	PR 05-MAR-1999; 99US-0123180.		
PR	25-AUG-1999;	99US-0150566.	PR 09-MAR-1999; 99US-0123548.		
PR	26-AUG-1999;	99US-0150884.	PR 23-MAR-1999; 99US-0125788.		
PR	27-AUG-1999;	99US-0151065.	PR 25-MAR-1999; 99US-0126264.		
PR	27-AUG-1999;	99US-0151066.	PR 29-MAR-1999; 99US-0126785.		
PR	27-AUG-1999;	99US-0151080.	PR 01-APR-1999; 99US-0127462.		
PR	30-AUG-1999;	99US-0151080.	PR 06-APR-1999; 99US-0128234.		
PR	31-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			



PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 22-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 26-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 52.4%; Score 43; DB 21; Length 376;  
 Best Local Similarity 61.5%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 PPLLEVGVEEKF 15  
 |||:|:|  
 Db 149 pplliqngqeenf 161

RESULT 11  
 ID AAY90913  
 AC AAY90913 standard; Protein; 3472 AA.  
 XX  
 AC AAY90913;  
 DT 30-AUG-2000 (first entry)  
 DE Cenarchaeum symbiosum open reading frame protein sequence SEQ ID NO:4.  
 XX  
 KW Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;  
 KW characterisation; archae; therapeutic; industrial; laboratory.  
 XX  
 OS Cenarchaeum symbiosum.  
 XX  
 PN WO200018909-A2.  
 XX  
 PD 06-APR-2000.  
 XX  
 PF 29-SEP-1999; 99WO-US22752.  
 XX  
 PR 29-SEP-1998; 98US-0102294.  
 XX  
 XX (DIVE-) DIVERSA CORP.  
 PA Swanson RV, Feldman RA, Schleper C;  
 XX WPI; 2000-293148/25.  
 DR N-PSDB; AAA55188.  
 DR  
 XX New nucleic acids and proteins isolated from the non-thermophilic  
 PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the  
 PT physiology of these archae and in therapeutic, industrial or laboratory  
 PT techniques -  
 XX  
 PS Claim 26; Page 102-110; 210pp; English.  
 XX

CC AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids  
 CC and proteins isolated from the non-thermophilic crenarchaeote  
 CC Cenarchaeum symbiosum. The nucleic acids and proteins identified in  
 CC the present invention are useful in characterising the physiology of  
 CC these archae and can be used in therapeutic, industrial or laboratory  
 CC techniques. AAA55227 to AAA55260 represent promoter sequences from  
 CC Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and  
 CC probes used in examples from the present invention.  
 XX  
 SQ Sequence 3472 AA;

Query Match 50.0%; Score 41; DB 21; Length 3472;  
 Best Local Similarity 66.7%; Pred. NO. 3.6e+02;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 VPPLLEVGVEE 13  
 :|||:|  
 Db 1169 ipplflvggae 1180

RESULT 12  
 ID AAY11325  
 AC AAY11325 standard; Protein; 63 AA.  
 XX  
 AC AAY11325;  
 XX  
 DT 20-MAY-1999 (first entry)  
 DE S. pneumoniae protein SEQ ID NO:435.  
 DE  
 XX S. pneumoniae pneumoniae strain 0100993; vaccine; immune response;  
 KW Streptococcus pneumoniae infection; pneumococcal.  
 KW Streptococcus pneumoniae.  
 OS  
 XX WO9737026-A1.  
 PN  
 PD 09-OCT-1997.  
 XX  
 PF 01-APR-1997; 97WO-US05306.  
 XX  
 PR 22-AUG-1996; 96US-0025788.  
 PR 02-APR-1996; 96US-0014690.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
 PI Stodola RK;  
 XX  
 DR WPI; 1997-503111/46.  
 DR N-PSDB; AAX30907.  
 XX  
 PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
 PT vaccines, drug screening, etc  
 XX  
 PS Claim 6; Page 320-321; 354pp; English.  
 XX  
 CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from  
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
 CC encode the novel proteins given in AAY11114 to AAY11367. The proteins,  
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
 CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.  
 XX  
 SQ Sequence 63 AA;

Query Match 48.8%; Score 40; DB 18; Length 63;

Best Local Similarity 42.9%; Pred. No. 7.1;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLLEVGVEEK 14  
|:|:|:| | : |  
Db 36 yippvvlndvdpnk 49

RESULT 13  
AAAY70541  
ID AAY70541 standard; Protein; 163 AA.  
XX  
AC AAY70541;  
XX  
DT 04-JUL-2000 (first entry)  
XX  
DE Protein encoded by N. tabacum Enh2 gene for enhancing INF1-induced HR.  
XX  
KW Enh2; enhancer of INF1-induced hypersensitive response; tobacco;  
KW resistance gene; R gene; non-host disease resistance; plant pathogen;  
KW hypersensitive response; HR; Phytophthora infestans elicitor; INF1;  
KW transgenic plant; viral pathogen; fungal pathogen; bacterial pathogen;  
KW nematodal pathogen; Phytophthora infestans.  
XX  
OS Nicotiana tabacum.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 159  
FT /note= "Encoded by tag"  
XX  
PN WO200012736-A2.  
XX  
PD 09-MAR-2000.  
XX  
PF 31-AUG-1999; 99WO-US19899.  
XX  
PR 31-AUG-1998; 98US-0098402.  
XX  
PA (MONS ) MONSANTO CO.  
XX  
PI Rommens CMT, Swords KMM, Yan H, Zhang B;  
XX  
DR WPI; 2000-256651/22.  
DR N-PSDB; AAZ51874.  
XX  
PT Identification of non-host plant disease resistance genes comprises  
PT expressing resistance and non-host inducible genes in susceptible  
PT plants .  
XX  
PS Claim 15; Page 69; 94pp; English.  
XX  
CC The patent discloses a method for identifying genes that enhance levels  
CC of disease resistance if expressed in susceptible plants. The method is  
CC useful for isolating disease resistance genes (R gene) in plants. These  
CC genes confer non-host disease resistance to plants by responding to  
CC avirulence genes in plant pathogens. The R-genes identified trigger a  
CC hypersensitive response (HR) in tobacco that is dependent on the presence  
CC of the Phytophthora infestans elicitor INF1. The genes are useful for  
CC generating pathogen-resistant transgenic plants. The transgenic plants  
CC are preferably Acacia, apple, banana, barley, bean, broccoli, cabbage,  
CC canola, carrot, citrus, coffee, corn, cotton, cucumber, Douglas fir,  
CC Eucalyptus, garlic, grape, Loblolly pine, melon, oat, oil palm, onion, an  
CC ornamental plant, pea, peanut, pepper, Poplar tree, potato, Radiata pine,  
CC rice, rye, sorghum, Southern pine, soybean, strawberry, sugarbeet,  
CC sugarcane, sunflower, Sweetgum, tea, tomato, turf, a vine and wheat. They  
CC can be used to control viral, fungal, bacterial or nematodal pathogens,  
CC e.g. Phytophthora, Erisyphe and Puccinia. The present sequence is encoded  
CC by partial sequence of Enh2 (enhancer of INF1-induced HR) tobacco R gene  
CC that enhances the INF1-dependent HR in Nicotiana benthamiana.  
XX  
SQ Sequence 163 AA;

Query Match 48.8%; Score 40; DB 21; Length 163;  
Best Local Similarity 58.3%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLLEVGVEEK 14  
|| ||:| | : :  
Db 149 ppellldvgfezq 160

RESULT 14  
AAAY86095  
ID AAY86095 standard; Protein; 172 AA.  
XX  
AC AAY86095;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE S. pneumoniae derived protein #304.  
XX  
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;  
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9806734-A1.  
XX  
PD 19-FEB-1998.  
XX  
PF 15-AUG-1997; 97WO-US14436.  
XX  
PR 16-AUG-1996; 96US-0024022.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
PI Stodola RK;  
XX  
DR WPI; 1998-159452/14.  
DR N-PSDB; AAZ96414.  
XX  
PT Streptococcus pneumoniae proteins and related DNA - useful for  
PT screening compounds for antibacterial activity  
XX  
PS Claim 5; Page 571; 640pp; English.  
XX  
CC This invention describes novel isolated Streptococcus pneumoniae  
CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see  
CC AAY85792-Y86182). The DNA, vectors and host cells described in the  
CC method of the invention are useful for the recombinant expression of the  
CC polypeptides. The polypeptides are useful for treatment or prevention of  
CC disease, or diagnosis of disease related to expression or activity of  
CC such a polypeptide. They can also be used to screen for compounds which  
CC interact with and inhibit or activate such a polypeptide. The  
CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
CC for inducing an immunological response in a mammal. The antagonists are  
CC useful to inhibit such bacterial polypeptides. The polypeptides are  
CC particularly useful to identify antimicrobial compounds and antibiotics.  
CC They are also useful to determine their role in pathogenesis of  
CC infection, dysfunction and disease.  
XX  
SQ Sequence 172 AA;

Query Match 48.8%; Score 40; DB 19; Length 172;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLLEVGVEEK 14  
|:|:|:| | : |  
Db 145 yippvvlndvdpnk 158

RESULT 15

AA82839  
ID AAR82839 standard; Protein; 437 AA.  
XX  
AC AAR82839;  
XX  
DT 25-JAN-1996 (first entry)  
XX  
DE Sugar beet citrate synthase.  
XX  
KW Citrate synthase; flower formation.  
XX  
OS Beta vulgaris strain Zuchtlinie 5S 0026.  
XX  
PN WO9524487-A.  
XX  
PD 14-SEP-1995.  
XX  
PF 07-MAR-1995; 95WO-EP00859.  
XX  
PR 19-OCT-1994; 94DE-4438821.  
PR 09-MAR-1994; 94DE-4408629.  
PR 22-SEP-1994; 94DE-4435366.  
XX  
PA (AGRE ) HOECHST-SCHERING AGREVO GMBH.  
XX  
PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;  
PI Mueller-roeber B;  
XX  
DR WPI; 1995-328278/42.  
DR N-PSDB; AAT04200.  
XX  
PT DNA encoding plant citrate synthase - used to regulate flower formation,  
PT to improve storage of tubers, etc. and to reduce sprouting  
XX  
PS Disclosure; Page 57-60; 87pp; English.  
XX  
CC To identify a cDNA from sugar beet which codes for citrate  
CC synthase, a cDNA bank of leaf tissue from sugar beet was prepd.  
CC Plaques of this cDNA bank were screened using radioactive DNA  
CC probes which comprise a mixture of Solanum tuberosum citrate  
CC synthase cDNA (AAT04199) and Nicotian tabacum citrate synthase cDNA  
CC (see AAT04201). One of the clones was sequenced. The nt. sequence is  
CC given in AAT04200.  
XX  
SQ Sequence 437 AA;

Query Match 48.8%; Score 40; DB 16; Length 437;  
Best Local Similarity 77.8%; Pred. No. 57;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPPLLEVG 10  
Db 356 vppillelg 364

RESULT 16  
AAY20069  
ID AAY20069 standard; Protein; 695 AA.  
XX  
AC AAY20069;  
XX  
DT 19-JUL-1999 (first entry)  
XX  
DE B. burgdorferi antigenic protein, t679.aa.  
DE  
XX  
KW Antigenic protein; vaccine; Lyme disease; infection; detection.  
XX  
OS Borrelia burgdorferi.  
XX  
PN WO9859071-Al.  
XX  
PD 30-DEC-1998.

XX 18-JUN-1998; 98WO-US12718.  
PF  
XX 03-SEP-1997; 97US-0057483.  
PR 20-JUN-1997; 97US-0050359.  
PR 22-JUL-1997; 97US-0053344.  
PR 22-JUL-1997; 97US-0053377.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMMUNE INC.  
XX  
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX  
DR WPI; 1999-189980/16.  
DR N-PSDB; AAX61766.  
XX  
PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
PT caused by Borrelia, particularly Lyme disease  
XX  
PS Claim 12; Page 183; 275pp; English.  
XX  
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus.  
XX  
SQ Sequence 695 AA;

Query Match 48.8%; Score 40; DB 20; Length 695;  
Best Local Similarity 50.0%; Pred. No. 94;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKFM 16  
Db 51 yvaylfkkgifekfv 66

RESULT 17  
AAY20068  
ID AAY20068 standard; Protein; 718 AA.  
XX  
AC AAY20068;  
XX  
DT 19-JUL-1999 (first entry)  
XX  
DE B. burgdorferi antigenic protein, f679.aa.  
DE  
XX  
KW Antigenic protein; vaccine; Lyme disease; infection; detection.  
XX  
OS Borrelia burgdorferi.  
XX  
PN WO9859071-Al.  
XX  
PD 30-DEC-1998.  
XX  
PF 18-JUN-1998; 98WO-US12718.  
XX  
PR 03-SEP-1997; 97US-0057483.  
PR 20-JUN-1997; 97US-0050359.  
PR 22-JUL-1997; 97US-0053344.  
PR 22-JUL-1997; 97US-0053377.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PA (MEDI-) MEDIMMUNE INC.  
XX  
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX  
DR WPI; 1999-189980/16.



DR N-PSDB; AAX61765.  
XX  
PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
PT caused by Borrelia, particularly Lyme disease  
XX  
PS Claim 12; Page 183; 275pp; English.  
XX  
CC This sequence represents a Borrelia burgdorferi (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the Borrelia genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the Borrelia genus. The products can also  
CC be used for detection of members of the Borrelia genus.  
XX  
SQ Sequence 718 AA;  
  
Query Match 48.8%; Score 40; DB 20; Length 718;  
Best Local Similarity 50.0%; Pred. No. 98;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 YVPPLLEVGVEEKF 16  
Db 74 yvaylfkkgifekfv 89  
  
RESULT 18  
AAY79062  
ID AAY79062 standard; peptide; 101 AA.  
XX  
AC AAY79062;  
XX  
DT 12-JUN-2000 (first entry)  
XX  
DE Polysaccharide binding domain of human PP1R6.  
XX  
KW Protein phosphatase 1; PP1; glycogen targeting subunit; blood glucose;  
KW phosphorylase alpha; glycogen synthesis; hyperglycaemic disorder;  
KW type I diabetes; type II diabetes; polysaccharide binding domain; PP1R6.  
XX  
OS Homo sapiens.  
XX  
PN WO200012549-A1.  
XX  
PD 09-MAR-2000.  
XX  
PF 19-AUG-1999; 99WO-GB02761.  
XX  
PR 27-AUG-1998; 98GB-0018650.  
XX  
PA (MEDI-) MEDICAL RES COUNCIL.  
XX  
PI Cohen PTW, Armstrong CG, Doherty MJ;  
XX  
DR WPI; 2000-256587/22.  
XX  
PT Lowering blood sugar levels in the treatment of diabetes, using a  
PT compound that blocks interaction between phosphorylase alpha and  
PT protein phosphatase 1 glycogen-targetting subunit -  
XX  
PS Disclosure; Fig 6; 51pp; English.  
XX  
CC This sequence represents the amino acid sequence of the human  
CC polysaccharide targeting domain of the human protein phosphatase 1  
CC glycogen targeting subunit PP1R6. The invention relates to the  
CC medicinal use of a compound capable of blocking the interaction of  
CC phosphorylase alpha with GL, where the compound comprises the C-terminal  
CC peptide of rat GL (see AAY79064) or a fragment of it. When phosphorylase  
CC alpha binds to GL it potentially inhibits its glycogen synthase phosphatase  
CC activity and inhibits glycogen synthesis, this contributes to high blood  
CC glucose levels. The invention also relates to a method for identifying

CC the compound. The compound can be used to reduce the blood glucose level  
CC of a mammal, particularly a human, in hyperglycaemic disorders such as  
CC type I or type II diabetes.  
XX  
SQ Sequence 101 AA;  
  
Query Match 47.6%; Score 39; DB 21; Length 101;  
Best Local Similarity 57.1%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 2 VPPLLEVGVEEKF 15  
Db 66 vppfllelgsrvhf 79  
  
RESULT 19  
AAY87121  
ID AAY87121 standard; Protein; 643 AA.  
XX  
AC AAY87121;  
XX  
DT 09-MAY-2000 (first entry)  
XX  
DE Human secreted protein sequence SEQ ID NO:160.  
XX  
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
KW antiinflammatory; nootropic; neuroprotective; antiallergic; cancer;  
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;  
KW foetal deficiency; blood disorder; immune system disorder; arthritis;  
KW autoimmune disease; hepatic disease; renal disease; inflammation;  
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;  
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;  
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;  
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;  
KW metabolic disorder; food additive; preservative.  
XX  
OS Homo sapiens.  
XX  
PN WO200004140-A1.  
XX  
PD 27-JAN-2000.  
XX  
PF 14-JUL-1999; 99WO-US15849.  
XX  
PR 15-JUL-1998; 98US-0092921.  
PR 15-JUL-1998; 98US-0092922.  
PR 15-JUL-1998; 98US-0092956.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;  
PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;  
PI Mucenski M, Endress GA, Soppet DR;  
XX  
DR WPI; 2000-161128/14.  
DR N-PSDB; AAZ98074.  
XX  
PT New isolated human genes, useful for diagnosis and treatment of, e.g.  
PT cancers, neurological or blood disorders -  
XX  
PS Claim 11; Page 415-417; 494pp; English.  
XX  
CC The polynucleotide sequences given in AAZ98017 to AAZ98108 encode the  
CC human secreted proteins given in AAY87064 to AAY87223. Human secreted  
CC protein can have activities based on the tissues and cells the genes are  
CC expressed in. Examples of activities include: cytostatic;  
CC immunosuppressive; antiinflammatory; nootropic; neuroprotective; and  
CC antiallergic. The polynucleotides and their corresponding secreted  
CC polypeptides are useful for preventing, treating or ameliorating medical  
CC conditions, e.g. by protein or gene therapy. Also pathological conditions  
CC can be diagnosed by determining the amount of the new polypeptides in a  
CC sample or by determining the presence of mutations in the new

CC polynucleotides. Human secreted protein s and their polynucleotides can  
CC be used for developing products for the diagnosis or treatment of cancer,  
CC tumors, neurodegenerative disorders, developmental abnormalities and  
CC foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, hepatic and renal disease, inflammation,  
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,  
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,  
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,  
CC cardiovascular disorders, reproductive disorders, gastrointestinal  
CC disorders, respiratory disorders and metabolic disorders. The  
CC proteins or polynucleotides can also be used as food additives or  
CC preservatives. The proteins are also useful for identifying their  
CC binding partners. AA298008 to AA298016 and AA298063 are sequence used in  
CC the exemplification of the present invention.

XX Sequence 643 AA;

Query Match 47.6%; Score 39; DB 21; Length 643;  
Best Local Similarity 60.0%; Pred. NO. 1.3e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 VPPLLEVGVEKFM 16  
Db 270 vppllcnmg--ekfl 282  
|||||:| |||:

RESULT 20  
AA298008  
ID AA298008 standard; Protein; 674 AA.  
XX AA298008;  
AC AA298008;  
DT 09-MAY-2000 (first entry)  
DE Human secreted protein sequence SEQ ID NO:248.

XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
KW antiinflammatory; neurotropic; neuroprotective; antiallergic; cancer;  
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;  
KW foetal deficiency; blood disorder; immune system disorder; arthritis;  
KW autoimmune disease; hepatic disease; renal disease; inflammation;  
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;  
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;  
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;  
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;  
KW metabolic disorder; food additive; preservative.

XX Homo sapiens.

XX WO200004140-A1.

XX 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15849.

XX 15-JUL-1998; 98US-0092921.

XX 15-JUL-1998; 98US-0092922.

XX 15-JUL-1998; 98US-0092956.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;

XX Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;

XX Mucenski M, Endress GA, Soppet DR;

XX WPI; 2000-161128/14.

XX N-PSDB; AA298074.

XX New isolated human genes, useful for diagnosis and treatment of, e.g.

XX cancers, neurological or blood disorders -

XX Disclosure; Page 479-482; 494pp; English.

XX The polynucleotide sequences given in AA298017 to AA298108 encode the  
CC human secreted proteins given in AA298064 to AA2987223. Human secreted  
CC protein can have activities based on the tissues and cells the genes are  
CC expressed in. Examples of activities include: cytostatic;  
CC immunosuppressive; antiinflammatory; neurotropic; neuroprotective; and  
CC antiallergic. The polynucleotides and their corresponding secreted  
CC polypeptides are useful for preventing, treating or ameliorating medical  
CC conditions, e.g. by protein or gene therapy. Also pathological conditions  
CC can be diagnosed by determining the amount of the new polypeptides in a  
CC sample or by determining the presence of mutations in the new  
CC polynucleotides. Human secreted proteins and their polynucleotides can  
CC be used for developing products for the diagnosis or treatment of cancer,  
CC tumours, neurodegenerative disorders, developmental abnormalities and  
CC foetal deficiencies, blood disorders, diseases of the immune system,  
CC autoimmune diseases, hepatic and renal disease, inflammation,  
CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,  
CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,  
CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,  
CC cardiovascular disorders, reproductive disorders, gastrointestinal  
CC disorders, respiratory disorders and metabolic disorders. The  
CC proteins or polynucleotides can also be used as food additives or  
CC preservatives. The proteins are also useful for identifying their  
CC binding partners. AA298008 to AA298016 and AA298063 are sequence used in  
CC the exemplification of the present invention.

XX Sequence 674 AA;

Query Match 47.6%; Score 39; DB 21; Length 674;  
Best Local Similarity 60.0%; Pred. NO. 1.4e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 VPPLLEVGVEKFM 16  
Db 301 vppllcnmg--ekfl 313  
|||||:| |||:

RESULT 21  
AA298017  
ID AA298017 standard; Protein; 58 AA.

XX AA298017;

XX 14-FEB-2001 (first entry)

XX Human secreted protein encoded by gene 41 clone HSIDA39.

XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO2000058358-A1.

XX 05-OCT-2000.

XX 23-MAR-2000; 2000WO-US07725.

XX 26-MAR-1999; 99US-0126602.

XX 14-JAN-2000; 2000US-0176063.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-594640/56.

XX N-PSDB; AAC79037.

XX Fourty nine nucleic acid molecules encoding human secreted proteins,

PT useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases -

XX

PS Claim 11; Page 353; 367pp; English.

XX

CC Sequences AAB44335-B44382 represent the amino acid sequences of 49 human secreted proteins encoded by the genes AAC69084-C69119. The genes and proteins are useful for preventing, ameliorating or treating medical conditions, e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

XX

SQ Sequence 58 AA;

Query Match 46.3%; Score 38; DB 21; Length 58;

Best Local Similarity 81.8%; Pred. No. 15;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLLLEVGVEEKF 15  
| | | | | | | |

Db 12 lllevpvaekf 22

RESULT 22

AAB42741

ID AAB42741 standard; Protein; 119 AA.

XX

AC AAB42741;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF2505 polypeptide sequence SEQ ID NO:5010.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnerable; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US08621.

XX

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX

PA (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

PI

XX

DR WPI; 2000-602362/57.

DR N-PSDB; AAC76950.

XX

PT Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

PT

XX

PS Claim 11; Page 4188; 5507pp; English.

XX

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnerary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX

SQ Sequence 119 AA;

Query Match 46.3%; Score 38; DB 21; Length 119;

Best Local Similarity 57.1%; Pred. No. 31;

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPLLLEVGVEEKF 16  
| | | | | | | |

Db 71 pnlllpfgteekll 84

RESULT 23

AAB50540

ID AAB50540 standard; Protein; 298 AA.

XX

AC AAB50540;

XX

DT 20-MAR-2001 (first entry)

XX

DE B. lactofermentum pyrimidine biosynthetic system enzyme protein #1.

XX

KW Brevibacterium lactofermentum; coryneform; microbe; dihydroorotase; pyrimidine biosynthetic system; aspartate carbamoyltransferase; orotidylate decarboxylase.

KW

KW

XX

OS Brevibacterium lactofermentum.

XX

PN JP2000287688-A.

XX

PD 17-OCT-2000.

XX

PF 08-APR-1999; 99JP-0100989.

XX

PR 08-APR-1999; 99JP-0100989.

XX

PA (AJIN ) AJINOMOTO KK.

XX

DR WPI; 2001-065574/08.

DR N-PSDB; AAC90812.  
PT A pyrimidine biosynthetic system enzyme gene of a coryneform microbe  
PS Claim 1; Page 18-19; 25pp; Japanese.  
XX  
XX The present invention describes a pyrimidine biosynthetic system enzyme  
CC gene of a coryneform microbe e.g. Brevibacterium lactofermentum. Also  
CC described are: (1) a DNA sequence encoding a protein comprising a  
CC defined 298 amino acid residue (P1) (AAB50540), or a sequence in which  
CC at least one amino acid residue is replaced, deleted, inserted, added or  
CC reversed in P1, where the protein has aspartate carbamoyltransferase  
CC activity; (2) a DNA encoding a protein comprising a defined 397 amino  
CC acid sequence (P2) (AAB50541) or a sequence in which at least one amino  
CC acid residue is replaced, deleted, inserted, added or reversed in P2,  
CC where the protein has dihydroorotase activity; (3) a DNA encoding a  
CC protein comprising a defined 278 amino acid sequence (P3) (AAB50548) or a  
CC sequence in which at least one amino acid residue is replaced, deleted,  
CC inserted, added or reversed in P3, where the protein has orotidylate  
CC decarboxylase activity; (4) a coryneform microbe transformed by the DNA;  
CC and (5) the proteins P1-P3 and their variants encoding by the DNA. The  
CC gene can be used for breeding nucleic acid producing microbes. The  
CC present sequence represents a protein encoded by a chromosome DNA  
CC fragment from Brevibacterium lactofermentum (AAC90812), which is a  
CC pyrimidine biosynthetic system enzyme.  
XX Sequence 298 AA;  
SQ

Query Match 46.3%; Score 38; DB 22; Length 298;  
Best Local Similarity 70.0%; Pred. No. 84;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 3 PPLLLEVGVE 12  
Db 178 pptllpvgve 187

RESULT 24  
AAB80121  
ID AAB80121 standard; Protein; 312 AA.  
XX  
AC AAB80121;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:976.  
XX  
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200100843-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-IB00923.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.

PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 14-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.  
PR 31-AUG-1999; 99DE-1041394.  
PR 31-AUG-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX  
XX (BADI ) BASF AG.  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX WPI; 2001-137957/14.  
XX N-PSDB; AAF72240.  
XX  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
XX pathway proteins, useful for producing fine chemicals in  
XX microorganisms, including organic acids, nonproteinogenic amino acids,  
XX and purine and pyrimidine bases -  
XX  
XX Claim 20; Page 1496-1497; 1737pp; English.  
XX  
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
XX pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
XX MP nucleic acids are useful for the production of fine chemicals  
XX in microorganisms, including organic acids, nonproteinogenic amino  
XX acids, purine and pyrimidine bases; nucleosides, nucleotides, lipids,  
XX saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
XX compounds, vitamins, cofactors, polyketides and enzymes.  
XX  
XX Sequence 312 AA;  
SQ



Query Match 46.3%; Score 38; DB 22; Length 312;  
Best Local Similarity 70.0%; Pred. No. 89;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLEVGVE 12  
Db 192 pptllpigue 201  
|||:||||

RESULT 25  
AAR09304  
ID AAR09304 standard; protein; 119 AA.  
XX AC AAR09304;  
XX DT 27-FEB-1991 (first entry)  
XX DE Sequence deduced from rev gene of HIV 1-NDK.  
XX KW Human immunodeficiency virus; AIDS.  
XX OS HIV 1-NDK.  
FH Key Location/Qualifiers  
FT Region 1..30  
FT /label= rev ex2  
FT Region 31..119  
FT /label= rev ex3  
XX WO9013630-A.  
PN 15-NOV-1990.  
XX 02-MAY-1990; 90WO-FR00312.  
XX PR 03-MAY-1989; 89FR-0005914.  
XX PA (INRM ) INSERM INST NAT SANTE.  
XX PI Barre-Sinoussi F, Chermann JC, Devaux C, Rey F, Sire J;  
PI Spire B;  
XX WPI; 1990-361470/48.  
DR N-PSDB; AAQ06635.  
XX New HIV-NDK retrovirus and protein component - used in vaccines  
PT against immuno-deficiency disorders and in raising MABs for  
PT retro-virus detection in vivo.  
XX Disclosure; Fig 2; 37pp; French.  
CC The HIV NDK virus was isolated from peripheral blood lymphocytes of  
CC an AIDS patient. A genomic library was prepd. from DNA extracted  
CC from CEM cells infected with the virus. The library was screened  
CC with a pBT1 probe corresp. to a fragment from HIV 1. The virus is  
CC more cytopathic than other strains and is not inhibited by OKT4A.  
CC It has been deposited as CNCM I-857. The sequence can be used to  
CC express proteins useful for diagnosing the presence of NDK and  
CC related viruses and in vaccines against immunodeficiency diseases.  
CC See also AAR09301-5.  
XX Sequence 119 AA;  
SQ

Query Match 45.1%; Score 37; DB 11; Length 119;  
Best Local Similarity 63.6%; Pred. No. 47;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLEVGVEE 13  
Db 109 ppavlesgte 119  
|||:||||

RESULT 26  
AAR20504  
ID AAR20504 standard; Protein; 398 AA..  
XX AC AAR20504;  
XX DT 06-MAY-1992 (first entry)  
XX DE Mutant Human alpha-1-antichymotrypsin with Arg(358).  
XX KW ACT; unglycosylated; serine protease inhibitor; serpin.  
XX OS Homo sapiens.  
XX PN US5079336-A.  
XX PD 07-JAN-1992.  
XX PF 23-JUN-1989; 89US-0370704.  
XX PR 23-JUN-1989; 89US-0370704.  
XX PA (UYPE-) UNIV OF PENNSYLVANI.  
XX PI Rubin H, Wang ZM;  
XX DR WPI; 1992-041064/05.  
XX PT New alpha-1-anti-chymotrypsin produced by recombinant DNA  
PT technology - for treatment of septic shock, pancreatitis,  
PT coagulation disorders, skin inflammation, etc.  
XX Claim 5; Column 16; 15pp; English.  
XX CC The amino acid sequence of wild-type ACT was derived from a cDNA  
CC clone which contained the entire human ACT coding region. Mutations  
CC were introduced into the coding sequence using site-directed  
CC techniques. In this mutant ACT, the active site (amino acids 358-9)  
CC has been mutated such that wild-type Leu at position 358 is  
CC substituted by Arg. The mutant has antichymotrypsin activity and  
CC anti-trypsin and antithrombin activity and can be used for treating  
CC blood clotting diseases and pancreatitis.  
CC See also AAQ20797 and AAR20503-R20507.  
XX Sequence 398 AA;  
SQ

Query Match 45.1%; Score 37; DB 13; Length 398;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEEKF 15  
Db 302 illqlgieeaf 312  
|||:|:|

RESULT 27  
AAR20505  
ID AAR20505 standard; Protein; 398 AA.  
XX AC AAR20505;  
XX DT 06-MAY-1992 (first entry)  
XX DE Mutant Human alpha-1-antichymotrypsin with Met(358).  
XX KW ACT; unglycosylated; serine protease inhibitor; serpin.  
XX OS Homo sapiens.  
XX PN US5079336-A.  
XX PD 07-JAN-1992.

XX 23-JUN-1989; 89US-0370704.  
XX 23-JUN-1989; 89US-0370704.  
XX (UYPE-) UNIV OF PENNSYLVANI.  
XX Rubin H, Wang ZM;  
XX WPI; 1992-041064/05.  
XX New alpha-1-anti-chymotrypsin produced by recombinant DNA  
XX technology - for treatment of septic shock, pancreatitis,  
XX coagulation disorders, skin inflammation, etc.  
XX Claim 6; Column 16; 15pp; English.  
XX The amino acid sequence of wild-type ACT was derived from a cDNA  
XX clone which contained the entire human ACT coding region. Mutations  
XX were introduced into the coding sequence using site-directed  
XX techniques. In this mutant ACT, the active site (amino acids 358-9)  
XX has been mutated such that wild-type Leu at position 358 is  
XX substituted by Met. The mutant has antichymotrypsin activity and  
XX anti-trypsin and antithrombin activity and can be used for treating  
XX blood clotting diseases and pancreatitis.  
XX See also AAQ20797 and AAR20503-R20507.  
XX Sequence 398 AA;  
XX  
XX Query Match 45.1%; Score 37; DB 13; Length 398;  
XX Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
XX Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 5 LLELVGVVEKF 15  
XX :||:|:|:|  
XX Db 302 illqlgieeaf 312  
XX  
XX RESULT 28  
XX AAR20506  
XX ID AAR20506 standard; Protein: 398 AA.  
XX AC AAR20506;  
XX DT 06-MAY-1992 (first entry)  
XX DE Mutant Human alpha-1-antichymotrypsin.  
XX KW ACT; unglycosylated; serine protease inhibitor; serpin.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX FT Active-site 358..359  
XX PN US5079336-A.  
XX PD 07-JAN-1992.  
XX PF 23-JUN-1989; 89US-0370704.  
XX PR 23-JUN-1989; 89US-0370704.  
XX PA (UYPE-) UNIV OF PENNSYLVANI.  
XX PI Rubin H, Wang ZM;  
XX WPI; 1992-041064/05.  
XX New alpha-1-anti-chymotrypsin produced by recombinant DNA  
XX technology - for treatment of septic shock, pancreatitis,  
XX coagulation disorders, skin inflammation, etc.

XX Claim 7; Column 16; 15pp; English.  
XX The amino acid sequence of wild-type ACT was derived from a cDNA  
XX clone which contained the entire human ACT coding region. Mutations  
XX were introduced into the coding sequence using site-directed  
XX techniques. In this mutant ACT, amino acids Ala(349) and Ala(350)  
XX have been substituted by Gly-Thr as a result of introducing a KpnI  
XX restriction site into the coding sequence; the introduction of a  
XX MluI restriction site at codons 368 and 369 has resulted in the  
XX substitution of Thr for wild-type Val at amino acid position 368.  
XX The mutations to the coding sequence allow the removal of the  
XX region coding for the active site; synthetic cassettes can be  
XX inserted to create new analogues.  
XX See also AAQ20797 and AAR20503-R20507.  
XX Sequence 398 AA;  
XX  
XX Query Match 45.1%; Score 37; DB 13; Length 398;  
XX Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
XX Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 5 LLELVGVVEKF 15  
XX :||:|:|:|  
XX Db 302 illqlgieeaf 312  
XX  
XX RESULT 29  
XX AAR20507  
XX ID AAR20507 standard; Protein: 398 AA.  
XX AC AAR20507;  
XX DT 06-MAY-1992 (first entry)  
XX DE Mutant Human alpha-1-antichymotrypsin with Phe(358).  
XX KW ACT; unglycosylated; serine protease inhibitor; serpin.  
XX OS Homo sapiens.  
XX PN US5079336-A.  
XX PD 07-JAN-1992.  
XX PF 23-JUN-1989; 89US-0370704.  
XX PR 23-JUN-1989; 89US-0370704.  
XX PA (UYPE-) UNIV OF PENNSYLVANI.  
XX PI Rubin H, Wang ZM;  
XX WPI; 1992-041064/05.  
XX New alpha-1-anti-chymotrypsin produced by recombinant DNA  
XX technology - for treatment of septic shock, pancreatitis,  
XX coagulation disorders, skin inflammation, etc.  
XX Claim 8; Column 16; 15pp; English.  
XX The amino acid sequence of wild-type ACT was derived from a cDNA  
XX clone which contained the entire human ACT coding region. Mutations  
XX were introduced into the coding sequence using site-directed  
XX techniques. In this mutant ACT, the active site (amino acids 358-9)  
XX has been mutated such that wild-type Leu at position 358 is  
XX substituted by Phe. The mutant has antichymotrypsin activity and  
XX anti-trypsin and antithrombin activity and can be used for treating  
XX blood clotting diseases and pancreatitis.  
XX See also AAQ20797 and AAR20503-R20506.  
XX Sequence 398 AA;  
XX

Query Match 45.1%; Score 37; DB 13; Length 398;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEEKF 15  
:|::|:| |  
Db 302 illqlgieeaf 312

RESULT 30  
AAR83104  
ID AAR83104 standard; Protein; 401 AA.  
XX  
AC AAR83104;  
XX  
DT 10-JUN-1996 (first entry)  
XX  
DE Alpha-1-antichymotrypsin rMAS-ACT analogue.  
XX  
KW Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;  
KW chymase inhibitor; protein engineering; analogue; infection;  
KW reperfusion injury; blood clot; lung inflammation; therapeutic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1..3 /label= MAS  
FT /note= "N-terminal extension (replaces LCHP of  
FT AAR83101)"  
FT Protein 4..401 /note= "mature protein"  
FT WO9527053-A1.  
XX  
PD 12-OCT-1995.  
XX  
PF 29-APR-1994; 94WO-US04703.  
XX  
PR 31-MAR-1994; 94US-0221171.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Cooperman BS, Rubin H, Schechter N, Wang ZM;  
XX  
DR WPI; 1995-366158/47.  
XX  
PS Analogues of human alpha-1-anti-chymotrypsin with increased  
PT inhibitory activity - useful for treating lung inflammation etc.  
PT also related nucleic acid, vectors and transformed cells.  
XX  
PS Claim 8; Fig 1A-1E; 54pp; English.  
XX  
CC Analogues of wild-type alpha-1-antichymotrypsin (a1-ACT) (AAR83101)  
CC may be produced, by recombinant DNA methods, in which amino acids  
CC 356-361 of the mature protein are replaced by other residues (e.g.  
CC see AAR83102-03), and optionally contain a modified N-terminal  
CC extension (see also AAR83105-06). The N-terminal extensions M-A-S  
CC or A-S on the wild-type and analogue proteins enable direct  
CC expression of stable monomers. The wild-type protein and analogues  
CC are used as chymase inhibitors and in the treatment and prevention  
CC of blood clots, reperfusion injury and lung inflammation, the  
CC latter caused by acid inhalation (from stomach contents or smoke)  
CC or infection by a Gram-negative bacterium (e.g. Pseudomonas or  
CC Escherichia). The analogues are prepared by protein engineering  
CC techniques, and are 4-fold more efficient at inhibiting chymase  
XX than the wild-type protein.  
SQ Sequence 401 AA;

Query Match 45.1%; Score 37; DB 16; Length 401;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEEKF 15  
:|::|:| |  
Db 305 illqlgieeaf 315

RESULT 31  
AAR83105  
ID AAR83105 standard; Protein; 401 AA.  
XX  
AC AAR83105;  
XX  
DT 10-JUN-1996 (first entry)  
XX  
DE Alpha-1-antichymotrypsin rMAS-ACT analogue.  
XX  
KW Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;  
KW chymase inhibitor; protein engineering; analogue; infection;  
KW reperfusion injury; blood clot; lung inflammation; therapeutic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1..3 /label= MAS  
FT /note= "N-terminal extension (replaces LCHP of  
FT AAR83102)"  
FT Protein 4..401 /note= "mature protein"  
FT Misc-difference 361 /label= L358X  
FT /note= "wild-type Leu-358 substitution"  
FT WO9527053-A1.  
XX  
PD 12-OCT-1995.  
XX  
PF 29-APR-1994; 94WO-US04703.  
XX  
PR 31-MAR-1994; 94US-0221171.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Cooperman BS, Rubin H, Schechter N, Wang ZM;  
XX  
DR WPI; 1995-366158/47.  
XX  
CC Analogues of human alpha-1-anti-chymotrypsin with increased  
CC inhibitory activity - useful for treating lung inflammation etc.  
CC also related nucleic acid, vectors and transformed cells.  
XX  
PS Claim 8; Fig 1A-1E; 54pp; English.  
XX  
CC Analogues of wild-type alpha-1-antichymotrypsin (a1-ACT) (AAR83101)  
CC may be produced, by recombinant DNA methods, in which amino acids  
CC 356-361 of the mature protein are replaced by other residues (e.g.  
CC see AAR83102-03), and optionally contain a modified N-terminal  
CC extension (see also AAR83104-06). The N-terminal extensions M-A-S  
CC or A-S on the wild-type and analogue proteins enable direct  
CC expression of stable monomers. The wild-type protein and analogues  
CC are used as chymase inhibitors and in the treatment and prevention  
CC of blood clots, reperfusion injury and lung inflammation, the  
CC latter caused by acid inhalation (from stomach contents or smoke)  
CC or infection by a Gram-negative bacterium (e.g. Pseudomonas or  
CC Escherichia). The analogues are prepared by protein engineering

CC techniques, and are 4-fold more efficient at inhibiting chymase  
CC than the wild-type protein.  
XX  
SQ Sequence 401 AA;

Query Match 45.1%; Score 37; DB 16; Length 401;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLEVGVEEKF 15  
:|:|:|:|:|  
Db 305 illqlgieeaf 315

RESULT 32  
AAR83106  
ID AAR83106 standard; Protein; 401 AA.  
XX  
AC AAR83106;

DT 10-JUN-1996 (first entry)  
XX  
DE Alpha-1-antichymotrypsin rMAS-ACT analogue.

XX Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;  
KW chymase inhibitor; protein engineering; analogue; infection;  
KW reperfusion injury; blood clot; lung inflammation; therapeutic.  
XX  
OS Synthetic.

XX Key Location/Qualifiers  
FH Misc-difference 1..3 /label= MAS  
FT /note= "N-terminal extension (replaces LCHP of  
FT AAR83103)"

FT Protein 4..401 /note= "mature protein"

FT Misc-difference 359..364 /label= IPXSIP

FT Misc-difference 361 /note= "replaces wild-type TLISAL"

FT /label= Met, Trp, Ala, Asn, Asp, Cys, Gln, Glu,  
FT Gly, His, Ile, Lys, Phe, Pro, Ser, Thr,  
FT Tyr, Val  
FT /note= "pref. Met (claim 3)"

XX WO9527053-A1.

XX 12-OCT-1995.

XX 29-APR-1994; 94WO-US04703.

XX 31-MAR-1994; 94US-0221171.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Cooperman BS, Rubin H, Schechter N, Wang ZM;

XX WPI; 1995-366158/47.

XX Analogues of human alpha-1-anti-chymotrypsin with increased  
PT inhibitory activity - useful for treating lung inflammation etc.  
PT also related nucleic acid, vectors and transformed cells.

XX Claim 8; Fig 1A-1E; 54pp; English.

XX Analogues of wild-type alpha-1-antichymotrypsin (a1-ACT) (AAR83101)  
CC may be produced, by recombinant DNA methods, in which amino acids  
CC 356-361 of the mature protein are replaced by other residues (e.g.  
CC see AAR83102-03), and optionally contain a modified N-terminal  
CC extension (see also AAR83104-05). The N-terminal extensions M-A-S  
CC or A-S on the wild-type and analogue proteins enable direct

CC expression of stable monomers. The wild-type protein and analogues  
CC are used as chymase inhibitors and in the treatment and prevention  
CC of blood clots, reperfusion injury and lung inflammation, the  
CC latter caused by acid inhalation (from stomach contents or smoke)  
CC or infection by a Gram-negative bacterium (e.g. Pseudomonas or  
CC Escherichia). The analogues are prepared by protein engineering  
CC techniques, and are 4-fold more efficient at inhibiting chymase  
XX than the wild-type protein.  
SQ Sequence 401 AA;

Query Match 45.1%; Score 37; DB 16; Length 401;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLEVGVEEKF 15  
:|:|:|:|:|  
Db 305 illqlgieeaf 315

RESULT 33  
AAR44435  
ID AAR44435 standard; Protein; 402 AA.

XX AAR44435;

XX 08-JUN-1994 (first entry)

XX Alpha-antichymotrypsin.

XX Alpha-antichymotrypsin; recombinant; spetic shock; coagulation;  
KW disorders; liver disease; skin; inflammation.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 4..398

XX Active-site /label= mat\_protein

XX US5266465-A.

XX 30-NOV-1993.

XX 23-JUN-1989; 89US-0370704.

XX 23-JUN-1989; 89US-0370704.

XX 24-JUL-1991; 91US-0735322.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Cooperman BS, Rubin H, Schechter N, Wang ZM;

XX WPI; 1993-395337/49.

XX P-PSDB; AAR44435.

XX Prodn. of analogues of alpha-1-anti-chymotrypsin useful in  
PT treating septic shock, skin inflammation etc., - by recombinant  
PT methods in which most cells transformed with DNA encoding  
PT analogue are cultured.

XX Disclosure; Fig 1a-le; 15pp; English.

XX The sequence is of an alpha-1-antichymotrypsin. The protein  
CC can be used to treat septic shock, blood clotting disorders,  
CC pancreatitis and liver diseases. The gene was manufactured by site  
CC directed mutagenesis using a commercially available kit.

XX Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;



Best Local Similarity 54.5%; Pred. No. 1.7e+02; Mismatches 4; Conservative 6; Indels 1; Gaps 0;

QY 5 LLELVGVVEEKF 15  
:|::|:| |  
Db 306 illqlgieeaf 316

RESULT 34  
AAR42737  
ID AAR42737 standard; Protein; 402 AA.  
XX  
AC AAR42737;  
XX  
DT 05-MAY-1994 (first entry)  
XX  
DE Alpha-1-antichymotrypsin (Leu358Arg).  
XX  
KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;  
KW pancreatitis; coagulation disorder; liver disease; enzyme;  
KW skin; inflammation; anti-thrombin; anti-trypsin.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Protein 5..402  
FT /label= mat\_protein  
XX  
PN US5252725-A.  
XX  
PD 12-OCT-1993.  
XX  
PF 23-JUN-1989; 89US-0370704.  
XX  
PR 23-JUN-1989; 89US-0370704.  
PR 24-JUL-1991; 91US-0735335.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Rubin H, Wang ZM;  
XX  
DR WPI; 1993-336158/42.  
XX  
PT Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -  
PT used for treating e.g. septic shock, pancreatitis, coagulation  
PT disorders, microbial disease and skin inflammation  
XX  
PS Claim 1; Fig 1; 15pp; English.  
XX

CC An isolated nucleic acid sequence encoding an analogue of ACT is  
CC claimed in which (i) Leu358 is substd. with Arg, Phe or Met or (ii)  
CC AlaAla349-350 are substd. with GlyThr, ValArg368-369 are substd.  
CC with ThrArg and Leu358 is opt. substd. with Phe.  
CC ACT analogues having the Met or Phe substdn. have anti-chymotrypsin  
CC activity and can be used for the treatment of septic shock,  
CC pancreatitis, coagulation disorders as in liver diseases, certain  
CC diseases caused by microbes that penetrate the skin by elaborating  
CC a microbial chymotrypsin-like enzyme and skin inflammation.  
CC ACT analogues having the Arg substdn. have anti-thrombin and  
CC anti-trypsin activity and can be used for treating blood clotting  
CC diseases and pancreatitis.  
XX  
SQ Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 5 LLELVGVVEEKF 15  
:|::|:| |  
Db 306 illqlgieeaf 316

RESULT 34  
AAR42739  
ID AAR42739 standard; Protein; 402 AA.  
XX  
AC AAR42739;  
XX

RESULT 35  
AAR42738  
ID AAR42738 standard; Protein; 402 AA.  
XX  
AC AAR42738;  
XX  
DT 05-MAY-1994 (first entry)  
XX  
DE Alpha-1-antichymotrypsin (Leu358Phe).  
XX

KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;  
KW pancreatitis; coagulation disorder; liver disease; enzyme;  
KW skin; inflammation; anti-thrombin; anti-trypsin.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Protein 5..402  
FT /label= mat\_protein  
XX  
PN US5252725-A.  
XX  
PD 12-OCT-1993.  
XX  
PF 23-JUN-1989; 89US-0370704.  
XX  
PR 23-JUN-1989; 89US-0370704.  
PR 24-JUL-1991; 91US-0735335.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Rubin H, Wang ZM;  
XX  
DR WPI; 1993-336158/42.  
XX  
PT Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -  
PT used for treating e.g. septic shock, pancreatitis, coagulation  
PT disorders, microbial disease and skin inflammation  
XX  
PS Claim 4; Fig 1; 15pp; English.  
XX

CC An isolated nucleic acid sequence encoding an analogue of ACT is  
CC claimed in which (i) Leu358 is substd. with Arg, Phe or Met or (ii)  
CC AlaAla349-350 are substd. with GlyThr, ValArg368-369 are substd.  
CC with ThrArg and Leu358 is opt. substd. with Phe.  
CC ACT analogues having the Met or Phe substdn. have anti-chymotrypsin  
CC activity and can be used for the treatment of septic shock,  
CC pancreatitis, coagulation disorders as in liver diseases, certain  
CC diseases caused by microbes that penetrate the skin by elaborating  
CC a microbial chymotrypsin-like enzyme and skin inflammation.  
CC ACT analogues having the Arg substdn. have anti-thrombin and  
CC anti-trypsin activity and can be used for treating blood clotting  
CC diseases and pancreatitis.  
XX  
SQ Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 5 LLELVGVVEEKF 15  
:|::|:| |  
Db 306 illqlgieeaf 316

RESULT 36  
AAR42739  
ID AAR42739 standard; Protein; 402 AA.  
XX  
AC AAR42739;  
XX

DT 05-MAY-1994 (first entry)  
DE Alpha-1-antichymotrypsin (Leu358Met).  
XX  
KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;  
KW pancreatitis; coagulation disorder; liver disease; enzyme;  
KW skin; inflammation; anti-thrombin; anti-trypsin.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Protein 5..402  
FT /label= mat\_protein  
FT Misc-difference 362  
FT /note= "Leu opt. substd. with Phe"  
FT Misc-difference 353..354  
FT /note= "AlaAla(349-350) substd. with GlyThr"  
FT Misc-difference 372..373  
FT /note= "ValArg(368-369) substd. with ThrArg"  
XX  
PN US5252725-A.  
XX  
PD 12-OCT-1993.  
XX  
PF 23-JUN-1989; 89US-0370704.  
XX  
PR 23-JUN-1989; 89US-0370704.  
PR 24-JUL-1991; 91US-0735335.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Rubin H, Wang ZM;  
XX  
DR WPI; 1993-336158/42.  
XX  
XX Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -  
PT used for treating e.g. septic shock, pancreatitis, coagulation  
PT disorders, microbial disease and skin inflammation  
PT  
XX  
PS Claim 2; Fig 1; 15pp; English.

XX  
CC An isolated nucleic acid sequence encoding an analogue of ACT is  
CC claimed in which (i) Leu358 is substd. with Arg, Phe or Met or (ii)  
CC AlaAla349-350 are substd. with GlyThr, ValArg368-369 are substd.  
CC with ThrArg and Leu358 is opt. substd. with Phe.  
CC ACT analogues having the Met or Phe substd. have anti-chymotrypsin  
CC activity and can be used for the treatment of septic shock,  
CC pancreatitis, coagulation disorders as in liver diseases, certain  
CC diseases caused by microbes that penetrate the skin by elaborating  
CC a microbial chymotrypsin-like enzyme and skin inflammation.  
CC ACT analogues having the Arg substd. have anti-thrombin and  
CC anti-trypsin activity and can be used for treating blood clotting  
CC diseases and pancreatitis.  
XX  
SQ Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLELVGVVEEKF 15  
:|::|::|::|  
Db 306 illqlgieeaf 316

RESULT 37  
AAR42740  
ID AAR42740 standard; Protein; 402 AA.  
XX  
AC AAR42740;  
XX  
DT 05-MAY-1994 (first entry)  
XX  
DE Alpha-1-antichymotrypsin analogue.

XX  
KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;  
KW pancreatitis; coagulation disorder; liver disease; enzyme;  
KW skin; inflammation; anti-thrombin; anti-trypsin.

XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Protein 5..402  
FT /label= mat\_protein  
FT Misc-difference 362  
FT /note= "Leu opt. substd. with Phe"  
FT Misc-difference 353..354  
FT /note= "AlaAla(349-350) substd. with GlyThr"  
FT Misc-difference 372..373  
FT /note= "ValArg(368-369) substd. with ThrArg"  
XX  
PN US5252725-A.  
XX  
PD 12-OCT-1993.  
XX  
PF 23-JUN-1989; 89US-0370704.  
XX  
PR 23-JUN-1989; 89US-0370704.  
PR 24-JUL-1991; 91US-0735335.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Rubin H, Wang ZM;  
XX  
DR WPI; 1993-336158/42.  
XX  
XX Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -  
PT used for treating e.g. septic shock, pancreatitis, coagulation  
PT disorders, microbial disease and skin inflammation  
PT  
XX  
PS Claim 3-4; Fig 1; 15pp; English.

XX  
CC An isolated nucleic acid sequence encoding an analogue of ACT is  
CC claimed in which (i) Leu358 is substd. with Arg, Phe or Met or (ii)  
CC AlaAla349-350 are substd. with GlyThr, ValArg368-369 are substd.  
CC with ThrArg and Leu358 is opt. substd. with Phe.  
CC ACT analogues having the Met or Phe substd. have anti-chymotrypsin  
CC activity and can be used for the treatment of septic shock,  
CC pancreatitis, coagulation disorders as in liver diseases, certain  
CC diseases caused by microbes that penetrate the skin by elaborating  
CC a microbial chymotrypsin-like enzyme and skin inflammation.  
CC ACT analogues having the Arg substd. have anti-thrombin and  
CC anti-trypsin activity and can be used for treating blood clotting  
CC diseases and pancreatitis.  
XX  
SQ Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLELVGVVEEKF 15  
:|::|::|::|  
Db 306 illqlgieeaf 316

RESULT 38  
AAR83101  
ID AAR83101 standard; Protein; 402 AA.  
XX  
AC AAR83101;  
XX  
DT 10-JUN-1996 (first entry)  
XX  
DE Wild-type alpha-1-antichymotrypsin.

XX  
KW Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;  
KW chymase inhibitor; protein engineering; analogue; infection;  
KW reperfusion injury; blood clot; lung inflammation; therapeutic.

OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1..4 /label= LCHP  
FT /note= "N-terminal, opt. replaced with N-terminal extension (M)AS (see AAR83104)"  
FT  
FT Protein 5..402 /note= "mature protein"  
FT  
FT WO9527053-A1.  
XX  
XX 12-OCT-1995.  
PD  
XX 29-APR-1994; 94WO-US04703.  
PF  
XX 31-MAR-1994; 94US-0221171.  
PR  
XX (UYPE-) UNIV PENNSYLVANIA.  
PA  
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;  
PI  
XX WPI; 1995-366158/47.  
DR  
DR N-PSDB; AAT00427.  
XX  
XX Analogues of human alpha-1-anti-chymotrypsin with increased inhibitory activity - useful for treating lung inflammation etc. also related nucleic acid, vectors and transformed cells.  
PT  
PT Claim 17; Fig 1a-1e; 54pp; English.  
XX  
XX Wild-type alpha-1-antichymotrypsin (al-ACT) may be produced by recombinant DNA methods. Analogues of al-ACT (AAR83102-06) may be produced in which amino acids 356-361 of the mature protein are replaced by other residues, and optionally contain a modified N-terminal extension. The wild-type protein and analogues are used as chymase inhibitors and in the treatment and prevention of blood clots, reperfusion injury and lung inflammation, the latter caused by acid inhalation (from stomach contents or smoke) or infection by a Gram-negative bacterium (e.g. Pseudomonas or Escherichia). The analogues are prepared by protein engineering techniques, and are 4-fold more efficient at inhibiting chymase than the wild-type protein.  
XX  
SQ Sequence 402 AA;  
  
Query Match 45.1%; Score 37; DB 16; Length 402;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
  
QY 5 LLEVGVEEKF 15  
Db 306 illqlgieef 316  
  
RESULT 39  
AAR83102  
ID AAR83102 standard; Protein; 402 AA.  
XX  
AC AAR83102;  
XX 10-JUN-1996 (first entry)  
XX Alpha-1-antichymotrypsin analogue.  
DE Alpha-1-antichymotrypsin analogue.  
XX Alpha-1-antichymotrypsin analogue; serine protease inhibitor; serpin; chymase inhibitor; protein engineering; infection; reperfusion injury; blood clot; lung inflammation; therapeutic.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers

FT Misc-difference 1..4 /label= LCHP  
FT /note= "N-terminal, opt. replaced with N-terminal extension (M)AS (see AAR83105)"  
FT  
FT Protein 5..402 /note= "mature rACT-L358 analogue"  
FT  
FT Misc-difference 362 /label= Trp, Ala, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Lys, Met, Phe, Pro, Ser, Thr, Tyr, Val  
FT /note= "wild-type Leu-358 substitution"  
FT  
FT Misc-difference 362 /label= L358X  
FT /note= "Trp (claim 6) (see AAT00428)"  
XX  
PN WO9527053-A1.  
XX 12-OCT-1995.  
PD  
XX 29-APR-1994; 94WO-US04703.  
PF  
XX 31-MAR-1994; 94US-0221171.  
PR  
XX (UYPE-) UNIV PENNSYLVANIA.  
PA  
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;  
PI  
XX WPI; 1995-366158/47.  
DR  
XX Analogues of human alpha-1-anti-chymotrypsin with increased inhibitory activity - useful for treating lung inflammation etc. also related nucleic acid, vectors and transformed cells.  
PT  
PT Claim 1; Fig 1A-1E; 54pp; English.  
XX  
XX The sequence represent an analogue of wild-type alpha-1-antichymotrypsin (al-ACT, AAR83101) having an alteration at Leu358 as shown above. Other analogues may be produced in which the 356-361 AA sequence of the mature protein is replaced by another sequence (see AAR83103), optionally with a modified N-terminal extension (see AAR83104-06). The wild-type protein and analogues are used as chymase inhibitors and in the treatment and prevention of blood clots, reperfusion injury and lung inflammation, the latter caused by acid inhalation (from stomach contents or smoke) or infection by a Gram-negative bacterium (e.g. Pseudomonas or Escherichia). The analogues are prepared by protein engineering techniques, and are 4-fold more efficient at inhibiting chymase than the wild-type protein.  
XX  
SQ Sequence 402 AA;  
  
Query Match 45.1%; Score 37; DB 16; Length 402;  
Best Local Similarity 54.5%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
  
QY 5 LLEVGVEEKF 15  
Db 306 illqlgieef 316  
  
RESULT 40  
AAR83103  
ID AAR83103 standard; Protein; 402 AA.  
XX  
AC AAR83103;  
XX 10-JUN-1996 (first entry)  
DF Alpha-1-antichymotrypsin analogue.  
DE Alpha-1-antichymotrypsin analogue.  
XX  
KW Alpha-1-antichymotrypsin; serine protease inhibitor; serpin; chymase inhibitor; protein engineering; analogue; infection;

KW reperfusion injury; blood clot; lung inflammation; therapeutic.  
XX Synthetic.  
OS  
XX  
FH Key  
FT Misc-difference 1.4  
FT /label= LCHP  
FT /note= "N-terminal, opt. replaced with N-terminal  
FT extension (M)AS (see AAR83106)."  
FT Protein 5.402 "mature protein"  
FT /note= 360.365  
FT Misc-difference /label= IPXSIP  
FT /note= "replaces wild-type TLLSAL"  
FT Misc-difference 362  
FT /label= Met, Trp, Ala, Asn, Asp, Cys, Gln, Glu,  
FT Gly, His, Ile, Lys, Phe, Pro, Ser, Thr,  
FT Tyr, Val  
FT /note= "pref. Met (claim 3)"  
XX WO9527053-A1.  
XX  
XX  
XX 12-OCT-1995.  
XX  
XX 29-APR-1994; 94WO-US04703.  
XX  
XX 31-MAR-1994; 94US-0221171.  
XX (UYPE-) UNIV PENNSYLVANIA.  
XX  
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;  
XX WPI; 1995-366158/47.  
XX  
XX Analogues of human alpha-1-anti-chymotrypsin with increased  
XX inhibitory activity - useful for treating lung inflammation etc.  
XX also related nucleic acid, vectors and transformed cells.  
XX  
XX Claim 2; Fig 1A-1E; 54pp; English.  
XX  
XX The sequence represent an analogue of wild-type alpha-1-  
XX antichymotrypsin (al-ACT, AAR83101). Other analogues may be  
XX produced (see AAR83102), optionally with modified N-terminal  
XX extensions (see AAR83104-06). The wild-type protein and analogues  
XX are used as chymase inhibitors and in the treatment and prevention  
XX of blood clots, reperfusion injury and lung inflammation, the latter  
XX caused by acid inhalation (from stomach contents or smoke) or  
XX infection by a Gram-negative bacterium (e.g. Pseudomonas or  
XX Escherichia). The analogues are prepared by protein engineering  
XX techniques, and are 4-fold more efficient at inhibiting chymase than  
XX the wild-type protein.  
XX  
XX Sequence 402 AA;

XX Alpha-1-antichymotrypsin; ACT; antithrombin; antitrypsin;  
KW chymotrypsin-inhibitor; protease-inhibitor; blood clotting;  
KW pancreatitis.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 1.4  
FT Peptide  
FT /note= "amino acids at positions 1-4 appear  
FT in the precursor of the mature protein"  
XX  
XX US5367064-A.  
XX  
XX 22-NOV-1994.  
XX  
XX 23-JUN-1989; 89US-0370704.  
XX  
XX 23-JUN-1989; 89US-0370704.  
PR 24-JUL-1991; 91US-0735335.  
PR 15-JAN-1993; 93US-0005908.  
XX  
XX (UYPE-) UNIV PENNSYLVANIA.  
XX  
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;  
PI WPI; 1995-005897/01.  
XX N-PSDB; AAQ75288.  
XX  
XX Nucleic acid encoding alpha-1-anti-chymotrypsin - for production  
XX of recombinant ACT and analogues having anti-trypsin,  
XX antithrombin and anti-chymotrypsin activity, useful in the  
XX treatment of blood clotting and pancreatitis  
XX  
XX Disclosure; Fig. 1a-ld; 16pp; English.  
XX  
XX A human liver cDNA library in phage lambda-gt11 was screened with  
XX polyclonal antisera raised against Cl esterase-inhibitor. The DNA  
XX sequence of the insert of a positive clone contained the entire  
XX coding region of human alpha-antichymotrypsin.  
XX  
XX Sequence 402 AA;  
XX  
XX Query Match 45.1%; Score 37; DB 16; Length 402;  
XX Best Local Similarity 54.5%; Pred. NO. 1.7e+02;  
XX Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 5 LLEVGVEEKF 15  
Db :||:|:| |  
306 illqlgieef 316  
RESULT 42  
AAR20503  
ID AAR20503 standard; Protein; 476 AA.  
XX  
XX AC AAR20503;  
XX  
XX 06-MAY-1992 (first entry)  
XX  
XX Human alpha-1-antichymotrypsin.  
XX  
XX ACT; unglycosylated; serine protease inhibitor; serpin.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 5.402  
FT Protein /label= ACT  
FT Active-site 362.363  
XX  
XX US5079336-A.  
PN

Query Match 45.1%; Score 37; DB 16; Length 402;  
Best Local Similarity 54.5%; Pred. NO. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 5 LLEVGVEEKF 15  
Db :||:|:| |  
306 illqlgieef 316  
RESULT 41  
AAR67259  
ID AAR67259 standard; Protein; 402 AA.  
XX  
XX AC AAR67259;  
XX  
XX 03-JUL-1995 (first entry)  
XX  
XX Alpha-1-antichymotrypsin.  
DE





FT /label= either or both substd. with P  
FT /note= "claimed analogue(s)"  
FT Misc-difference 359  
FT /label= substd. with P  
FT /note= "claimed analogue"  
FT Misc-difference 360..361  
FT /label= either or both substd. with P  
FT /note= "claimed analogue(s)"  
FT Misc-difference 358  
FT /label= M,I,V,A,D,T,E  
FT /note= "claimed analogue(s)"  
XX  
XX WO9527055-A.  
XX  
XX 12-OCT-1995.  
XX  
XX 29-APR-1994; 94WO-US04735.  
XX  
XX 31-MAR-1994; 94US-0221078.  
XX  
XX (UYPE-) UNIV PENNSYLVANIA.  
XX  
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;  
XX  
XX WPI; 1995-358630/46.  
XX N-PSDB; AAT03876.  
XX  
XX Analogue of human alpha-1-anti-chymotrypsin - acts as an efficient  
XX inhibitor of human neutrophil elastase, for use in treating inflammation  
XX  
XX Disclosure; Fig 1; 51pp; English.

CC A human liver cDNA library was screened with polyclonal antisera  
CC raised against C1 esterase inhibitor. Positive clones were  
CC sequenced. The DNA sequence and the derived AA sequence of the Eco  
CC R1 fragment from one of the positive lambda-gt11 cDNA clones  
CC contained the entire coding region of the mature human alpha-1-  
CC antichymotrypsin (alpha AC), as depicted in AAT03876. The construct  
CC also included a 21 nt extension of the 5' end encoding 7 AAs,  
CC comprising the sequence in AAR82603. The mature protein has an Mr  
CC 45,031. AAs corresp. to AAR82601 at AA posns. 356-361 were substd.  
CC with AAR82602. AAR82602 was selected from the corresp. segment of  
CC alpha-1-proteinase inhibitor, i.e. the reactive center of alpha-1-  
CC proteinase inhibitor from the p3 through the p3' posns. The  
CC resulting analogue is claimed. Other analogues are also claimed  
CC (see AAR82250 FT). The analogues have neutrophil elastase  
CC inhibiting activity.

XX Sequence 476 AA;

Query Match 45.1%; Score 37; DB 16; Length 476;  
Best Local Similarity 54.5%; Pred. No. 2.le+02;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLEVGVEEKF 15  
Db 306 illqlgieef 316  
:::|:|:|

RESULT 45  
AAG31539  
ID AAG31539 standard; Protein; 1410 AA.  
XX  
XX AAG31539;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 37892.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX  
XX PD 06-SEP-2000.  
XX  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.  
XX PR 23-APR-1999; 99US-0130510.  
XX PR 23-APR-1999; 99US-0130891.  
XX PR 28-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 30-APR-1999; 99US-0132407.  
XX PR 04-MAY-1999; 99US-0132484.  
XX PR 05-MAY-1999; 99US-0132485.  
XX PR 06-MAY-1999; 99US-0132486.  
XX PR 07-MAY-1999; 99US-0132487.  
XX PR 07-MAY-1999; 99US-0132863.  
XX PR 11-MAY-1999; 99US-0134256.  
XX PR 14-MAY-1999; 99US-0134218.  
XX PR 14-MAY-1999; 99US-0134219.  
XX PR 14-MAY-1999; 99US-0134221.  
XX PR 14-MAY-1999; 99US-0134370.  
XX PR 18-MAY-1999; 99US-0134768.  
XX PR 19-MAY-1999; 99US-0134941.  
XX PR 20-MAY-1999; 99US-0135124.  
XX PR 21-MAY-1999; 99US-0135353.  
XX PR 24-MAY-1999; 99US-0135629.  
XX PR 25-MAY-1999; 99US-0136021.  
XX PR 27-MAY-1999; 99US-0136392.  
XX PR 28-MAY-1999; 99US-0136782.  
XX PR 01-JUN-1999; 99US-0137222.  
XX PR 03-JUN-1999; 99US-0137528.  
XX PR 04-JUN-1999; 99US-0137502.  
XX PR 07-JUN-1999; 99US-0137724.  
XX PR 08-JUN-1999; 99US-0138094.  
XX PR 10-JUN-1999; 99US-0138540.  
XX PR 14-JUN-1999; 99US-0138847.  
XX PR 16-JUN-1999; 99US-0139119.  
XX PR 16-JUN-1999; 99US-0139452.  
XX PR 16-JUN-1999; 99US-0139453.  
XX PR 17-JUN-1999; 99US-0139492.  
XX PR 18-JUN-1999; 99US-0139454.  
XX PR 18-JUN-1999; 99US-0139455.  
XX PR 18-JUN-1999; 99US-0139456.  
XX PR 18-JUN-1999; 99US-0139457.  
XX PR 18-JUN-1999; 99US-0139458.  
XX PR 18-JUN-1999; 99US-0139459.  
XX PR 18-JUN-1999; 99US-0139460.  
XX PR 18-JUN-1999; 99US-0139461.  
XX PR 18-JUN-1999; 99US-0139462.  
XX PR 18-JUN-1999; 99US-0139463.  
XX PR 18-JUN-1999; 99US-0139750.  
XX PR 21-JUN-1999; 99US-0139763.  
XX PR 21-JUN-1999; 99US-0139817.  
XX PR 22-JUN-1999; 99US-0139899.  
XX PR 23-JUN-1999; 99US-0140353.  
XX PR 23-JUN-1999; 99US-0140354.  
XX PR 24-JUN-1999; 99US-0140695.  
XX PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 45.1%; Score 37; DB 21; Length 1410;  
Best Local Similarity 63.6%; Pred. No. 6.7e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLLEVGVEE 13  
|||: : ||||  
Db 668 ppliekfgvee 678

Search completed: June 28, 2001, 11:54:12  
Job time: 73 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 37.02 Seconds  
(without alignments)  
8.706 Million cell updates/sec

Title: US-09-439-313-554  
Perfect score: 82  
Sequence: 1 YVPPLLLLEGVVEEKFM 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	48.8	51	1	US-08-346-849-15
2	40	48.8	51	2	US-08-293-284A-15
3	39	47.6	1544	4	US-09-413-814-46
4	36	43.9	415	1	US-08-381-936-2
5	36	43.9	415	3	US-08-943-374-2
6	35.5	43.3	525	2	US-08-749-902-7
7	35.5	43.3	525	2	US-08-749-902-8
8	35	42.7	32	3	US-08-938-548B-7
9	35	42.7	80	3	US-08-554-840-13
10	35	42.7	123	3	US-08-938-548B-10
11	35	42.7	130	3	US-08-938-548B-6
12	35	42.7	424	4	US-08-876-885-26
13	35	42.7	486	2	US-08-432-016-6
14	35	42.7	486	2	US-08-684-594-6
15	35	42.7	2938	5	PCT-US94-00198-3
16	34	41.5	55	4	US-09-042-012-15
17	34	41.5	63	4	US-09-042-012-9
18	34	41.5	63	4	US-09-042-012-11
19	34	41.5	64	4	US-09-042-012-7
20	34	41.5	64	4	US-09-305-086-1
21	34	41.5	71	4	US-09-042-012-17
22	34	41.5	89	3	US-08-946-026-1
23	34	41.5	260	4	US-08-081-929-10
24	34	41.5	271	2	US-08-937-972-6
25	34	41.5	315	1	US-08-129-129-8
26	34	41.5	321	2	US-08-937-972-3
27	34	41.5	494	3	US-08-993-260-3

28	34	41.5	858	3	US-08-946-026-3	Sequence 3, Appli
29	34	41.5	1311	2	US-08-540-406-4	Sequence 4, Appli
30	34	41.5	1311	3	US-08-656-055-4	Sequence 4, Appli
31	34	41.5	1311	4	US-08-954-668-4	Sequence 4, Appli
32	34	41.5	1311	5	PCT-US95-13233-4	Sequence 4, Appli
33	33.5	40.9	379	4	US-09-413-304-15	Sequence 15, Appl
34	33	40.2	125	4	US-08-905-223-320	Sequence 320, App
35	33	40.2	328	2	US-08-977-767-1	Sequence 1, Appli
36	33	40.2	340	2	US-08-974-546-5	Sequence 5, Appli
37	33	40.2	366	3	US-08-987-904A-2	Sequence 2, Appli
38	33	40.2	366	3	US-08-987-904A-4	Sequence 4, Appli
39	33	40.2	395	3	US-09-080-044-5	Sequence 5, Appli
40	33	40.2	459	4	US-09-080-983-5	Sequence 5, Appli
41	33	40.2	906	2	US-08-609-230A-9	Sequence 9, Appli
42	33	40.2	907	3	US-08-990-140-4	Sequence 4, Appli
43	33	40.2	940	4	US-08-810-712-7	Sequence 7, Appli
44	33	40.2	1299	4	US-08-460-900C-62	Sequence 62, Appl
45	33	40.2	1434	2	US-08-540-406-10	Sequence 10, Appl
46	33	40.2	1434	3	US-08-656-055-10	Sequence 10, Appl
47	33	40.2	1434	4	US-08-954-668-10	Sequence 10, Appl
48	33	40.2	1434	5	PCT-US95-13233-10	Sequence 10, Appl
49	33	40.2	1447	2	US-08-540-406-19	Sequence 19, Appl
50	33	40.2	1447	3	US-08-656-055-19	Sequence 19, Appl
51	33	40.2	1447	4	US-08-954-668-19	Sequence 19, Appl
52	33	40.2	1447	5	PCT-US95-13233-19	Sequence 19, Appl
53	33	40.2	2500	2	US-08-801-263A-2	Sequence 2, Appli
54	33	40.2	2500	3	US-09-102-248-2	Sequence 2, Appli
55	33	40.2	2512	2	US-08-801-263A-9	Sequence 9, Appli
56	33	40.2	2512	3	US-09-102-248-9	Sequence 9, Appli
57	33	40.2	2517	2	US-08-801-263A-5	Sequence 5, Appli
58	33	40.2	2517	3	US-09-102-248-5	Sequence 5, Appli
59	32	39.0	20	2	US-08-564-972-21	Sequence 21, Appl
60	32	39.0	193	2	US-08-564-972-8	Sequence 8, Appli
61	32	39.0	194	1	US-08-148-058A-27	Sequence 27, Appl
62	32	39.0	194	1	US-08-148-058A-29	Sequence 29, Appl
63	32	39.0	194	1	US-08-478-042-27	Sequence 27, Appl
64	32	39.0	194	1	US-08-478-042-29	Sequence 29, Appl
65	32	39.0	194	2	US-08-645-215-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1  
US-08-346-849-15  
; Sequence 15, Application US/08346849  
; Patent No. 5670483  
; GENERAL INFORMATION:  
; APPLICANT: Zhang, Shuguang  
; APPLICANT: Lockshin, Curtis  
; APPLICANT: Rich, Alexander  
; APPLICANT: Holmes, Todd  
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY  
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02173-4799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/346,849  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:



```

; ZIP: 10036-8403
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/381,936
; FILING DATE: 09-FEB-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93 02110
; FILING DATE: 09-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42270618
; FILING DATE: 08-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-381-936-2

```

Query Match 43.9%; Score 36; DB 1; Length 415;  
Best Local Similarity 46.2%; Pred. No. 94;  
Matches 6; Conservative 4; Mismatches 3; Indels

Qy 2 VPPLLEVGVEEK 14  
: : : : :  
Db 273 LPPLITAVGVND 285

## RESULT

US-08-943-374-2  
; Sequence 2, Application US/08943374  
; Patent No. 6028249  
; GENERAL INFORMATION:  
; APPLICANT: ROBER, Manuela  
; APPLICANT: GEIER, Gebhardt  
; APPLICANT: GEIDER, Klaus  
; APPLICANT: WILLMITZER, Lothar  
; TITLE OF INVENTION: DNA sequences which lead to the  
; TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing  
; TITLE OF INVENTION: these sequences as well as a process for preparing  
; TITLE OF INVENTION: transgenic plants.  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,374  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/381,936  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: DE P42270618  
 ; FILING DATE: 08-DEC-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meilman, Edward A.  
 ; REGISTRATION NUMBER: 24,735  
 ; REFERENCE/DOCKET NUMBER: P/951-108  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 382-0700  
 ; TELEFAX: (212) 382-0888  
 ; TELEX: 236925  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 415 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-943-374-2

Query Match 43.9%; Score 36; DB 3; Length 415;  
Best Local Similarity 46.2%; Pred. NO. 94;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEK 14  
:| | | : | | | :  
Db 273 LPPLITAVGVNDQ 285

## RESULT

```

US-08-749-902-7
; Sequence 7, Application US/08749902
; Patent No. 5985635
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN SERINE/THREONINE
; TITLE OF INVENTION: PROTEIN KINASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,902
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0150 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 189508  
US-08-749-902-7

Query Match 43.3% Score 35.5; DB 2; Length 525;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 YVPPLLEVGVEEKF 15  
||| :||| :|||  
Db 413 YVAPSVLE-SVKEKF 426

RESULT 7

US-08-749-902-8  
; Sequence 8, Application US/08749902  
; Patent No. 5985635  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: NOVEL HUMAN SERINE/THREONINE  
; TITLE OF INVENTION: PROTEIN KINASES  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/749,902  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0150 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 525 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1562  
US-08-749-902-8

Query Match 43.3% Score 35.5; DB 2; Length 525;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 YVPPLLEVGVEEKF 15  
||| :||| :|||  
Db 413 YVAPSVLE-SVKEKF 426

RESULT 8

US-08-938-548B-7  
; Sequence 7, Application US/08938548B  
; Patent No. 6001963  
; GENERAL INFORMATION:  
; APPLICANT: Yanagisawa, Masashi  
; APPLICANT: Bergsma, Derk  
; APPLICANT: Wilson, Shelagh  
; APPLICANT: Brooks, David  
; APPLICANT: Gellai, Miklos  
; TITLE OF INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE  
; TITLE OF INVENTION: RECEPTOR HFGAN72  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: United States of America  
; ZIP: 19406  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/938,548B  
; FILING DATE: 26-SEPT-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/887,382  
; FILING DATE: 2-JUL-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/820,519  
; FILING DATE: 19-MAR-1997  
; APPLICATION NUMBER: 60/033,604  
; FILING DATE: 17-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elizabeth J. Hecht  
; REGISTRATION NUMBER: 41,824  
; REFERENCE/DOCKET NUMBER: ATG50037-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5009  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-938-548B-7

Query Match 42.7% Score 35; DB 3; Length 32;  
Best Local Similarity 54.5%; Pred. No. 7;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12  
:||| :||| :|||  
Db 21 LPPALLSLGVD 31

RESULT 9

US-08-554-840-13  
; Sequence 13, Application US/08554840  
; Patent No. 6001358  
; GENERAL INFORMATION:  
; APPLICANT: BLACK, Amelia  
; APPLICANT: HANNA, Nabil  
; APPLICANT: PADLAN, Eduardo A.  
; APPLICANT: NEWMAN, Roland A.  
; TITLE OF INVENTION: HUMANIZED ANTIBODIES TO HUMAN gp39,



; TITLE OF INVENTION: COMPOSITIONS CONTAINING AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/554,840
; FILING DATE: 07-NOV-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-554-840-13

Query Match 42.7%; Score 35; DB 3; Length 80;
Best Local Similarity 53.8%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 PPLLLEVGVEKF 15
|| || || || ||
Db 32 PKLLIYGVDPDRF 44

RESULT 10
US-08-938-548B-10
; Sequence 10, Application US/08938548B
; Patent No. 6001963
; GENERAL INFORMATION:
; APPLICANT: Yanagisawa, Masashi
; APPLICANT: Bergsma, Derk
; APPLICANT: Wilson, Shelagh
; APPLICANT: Brooks, David
; APPLICANT: Gellai, Miklos
; TITLE OF INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: United States of America
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,548B
; FILING DATE: 26-SEPT-1997
; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/887,382
; FILING DATE: 2-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,519
; FILING DATE: 19-MAR-1997
; APPLICATION NUMBER: 60/033,604
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth J. Hecht
; REGISTRATION NUMBER: 41,824
; REFERENCE/DOCKET NUMBER: ATG50037-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5009
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 123 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-938-548B-10

Query Match 42.7%; Score 35; DB 3; Length 123;
Best Local Similarity 54.5%; Pred. No. 34;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12
: || || || ||
Db 14 LPPALLSLGVD 24

RESULT 11
US-08-938-548B-6
; Sequence 6, Application US/08938548B
; Patent No. 6001963
; GENERAL INFORMATION:
; APPLICANT: Yanagisawa, Masashi
; APPLICANT: Bergsma, Derk
; APPLICANT: Wilson, Shelagh
; APPLICANT: Brooks, David
; APPLICANT: Gellai, Miklos
; TITLE OF INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: United States of America
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,548B
; FILING DATE: 26-SEPT-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/887,382
; FILING DATE: 2-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,519
; FILING DATE: 19-MAR-1997
; APPLICATION NUMBER: 60/033,604
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:

NAME: Elizabeth J. Hecht  
 REGISTRATION NUMBER: 41,824  
 REFERENCE/DOCKET NUMBER: ATG50037-2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 610-270-5009  
 TELEFAX: 610-270-5090  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 130 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-938-548B-6

Query Match 42.7%; Score 35; DB 3; Length 130;  
 Best Local Similarity 54.5%; Pred. No. 36;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12  
 Db 21 LPPALLSLGVD 31

RESULT 12  
 US-08-876-885-26  
 ; Sequence 26, Application US/08876885  
 ; Patent No. 6174713  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shen, Xiaoyu  
 ; APPLICANT: Houtan, Fariba  
 ; TITLE OF INVENTION: CANDIDA CYTOPLASMIC TRYPTOPHANYL-trNA  
 ; TITLE OF INVENTION: SYNTHETASE PROTEINS, NUCLEIC ACIDS AND STRAINS COMPRISING  
 ; TITLE OF INVENTION: SAME  
 ; NUMBER OF SEQUENCES: 26  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/876,885  
 ; FILING DATE: 16-JUN-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brook, David E.  
 ; REGISTRATION NUMBER: 22,592  
 ; REFERENCE/DOCKET NUMBER: CPI97-02  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781) 861-6240  
 ; TELEFAX: (781) 861-9540  
 ; INFORMATION FOR SEQ ID NO: 26:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 424 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; US-08-876-885-26

Query Match 42.7%; Score 35; DB 4; Length 424;  
 Best Local Similarity 46.2%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 4 PLLLEVGVEKEM 16  
 Db 129 PLVIELTDEKFL 141  
 RESULT 13  
 US-08-432-016-6  
 ; Sequence 6, Application US/08432016  
 ; Patent No. 5968768  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HAYNES, BARTON F.  
 ; APPLICANT: APUFFO, ALEJANDRO  
 ; APPLICANT: PATEL, DHAVALKUMAR  
 ; APPLICANT: BOWEN, MICHAEL A.  
 ; APPLICANT: MARQUARDT, HANS  
 ; TITLE OF INVENTION: CD6 LIGAND  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON & VANDERHYE P.C.  
 ; STREET: 1100 NORTH GLEBE ROAD  
 ; CITY: ARLINGTON  
 ; STATE: VIRGINIA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22201-4714  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/432,016  
 ; FILING DATE: 01-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/333,350  
 ; FILING DATE: 02-NOV-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/143,903  
 ; FILING DATE: 02-NOV-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: WILSON, MARY J.  
 ; REGISTRATION NUMBER: 32,955  
 ; REFERENCE/DOCKET NUMBER: 1579-95  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 816-4000  
 ; TELEFAX: (703) 816-4100  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 486 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-432-016-6

Query Match 42.7%; Score 35; DB 2; Length 486;  
 Best Local Similarity 72.7%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12  
 Db 448 VTPELLETGVE 458

RESULT 14  
 US-08-684-594-6  
 ; Sequence 6, Application US/08684594  
 ; Patent No. 5998172  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HAYNES, BARTON F.  
 ; APPLICANT: APUFFO, ALEJANDRO  
 ; APPLICANT: PATEL, DHAVALKUMAR

```
; APPLICANT: BOWEN, MICHAEL A.
; APPLICANT: MARQUARDT, HANS
; TITLE OF INVENTION: CD6 LIGAND
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,594
; FILING DATE: 18-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,016
; FILING DATE: 01-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/333,350
; FILING DATE: 02-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/143,903
; FILING DATE: 02-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-684-594-6

Query Match 42.7%; Score 35; DB 2; Length 486;
Best Local Similarity 72.7%; Pred. No. 1.7e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12
Db 448 VTPELLETGVE 458

RESULT 15
PCT-US94-00198-3
; Sequence 3, Application PC/TUS9400198
; GENERAL INFORMATION:
; APPLICANT: Schering Corp.
; TITLE OF INVENTION: RAS Associated GAP Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering Corp.
; STREET: 1 Giralld Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 6.0.8
```

```
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00198
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/004,824
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: DX0352 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)822-7255
; TELEFAX: (201)822-7039
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2938 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; PCT-US94-00198-3

Query Match 42.7%; Score 35; DB 5; Length 2938;
Best Local Similarity 58.3%; Pred. NO. 1.4e+03;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLEVGVEEKF 16
Db 2762 LLEVGFEDEL 2773

RESULT 16
US-09-042-012-15
; Sequence 15, Application US/09042012A
; Patent No. 6111087
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
; APPLICANT: WINTER, Jan
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-006
; CURRENT APPLICATION NUMBER: US/09/042,012A
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 55
; TYPE: PRT
; ORGANISM: transmembrane anchor domain of HFV env protein
; US-09-042-012-15

Query Match 41.5%; Score 34; DB 4; Length 55;
Best Local Similarity 54.5%; Pred. NO. 20;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVPPLLEVG 11
Db 24 YLKPILIGVG 34

RESULT 17
US-09-042-012-9
; Sequence 9, Application US/09042012A
; Patent No. 6111087
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
```





```

; Sequence 1, Application US/08946026
; Patent No. 6034218
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Mitcham, Jennifer L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND IMMUNODIAGNOSIS OF PROSTATE CANCER
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,026
; FILING DATE: 07-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 89 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-946-026-1

```

```

Query Match 41.5%; Score 34; DB 3; Length 89;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 6 LLEVGVEKF 15
|: |:|:|
Db 26 LMSIGLDEKF 35

```

```

RESULT 23
US-08-081-929-10
; Sequence 10, Application US/08081929
; Patent No. 6160090
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; APPLICANT: Barnea, Gilad
; APPLICANT: Grumet, Martin H.
; APPLICANT: Margolis, Richard U.
; TITLE OF INVENTION: A NEW CLASS OF RPTPasases: THEIR
; TITLE OF INVENTION: STRUCTURAL DOMAINS AND LIGANDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/081,929
; FILING DATE: 23-JUN-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-081-929-10

```

```

Query Match 41.5%; Score 34; DB 4; Length 260;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 5 LLEVGVEE 13
|: |:|:|
Db 151 ILFEIGVEE 159

```

```

RESULT 24
US-08-937-972-6
; Sequence 6, Application US/08937972
; Patent No. 5932443
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: ANTIGENS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,972
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0400 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 271 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 199586  
US-08-937-972-6

Query Match 41.5%; Score 34; DB 2; Length 271;  
Best Local Similarity 46.2%; Pred. NO. 1.3e+02;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 VPPLLEVGVEEK 14  
Db 237 VEPILSVGEDE 249

RESULT 25  
US-08-129-129-8  
; Sequence 8, Application US/08129129  
; Patent No. 5767363  
; GENERAL INFORMATION:  
; APPLICANT: DE SILVA, Jacqueline  
; APPLICANT: SAFFORD, Richard  
; APPLICANT: HUGHES, Stephen Glyn  
; TITLE OF INVENTION: PLANT PROMOTER INVOLVED IN  
; TITLE OF INVENTION: CONTROLLING LIPID BIOSYNTHESIS IN SEEDS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARB & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/129,129  
; FILING DATE: 18-JAN-1994  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91303098.7  
; FILING DATE: 09-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB92/00627  
; FILING DATE: 08-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, Paul N.  
; REGISTRATION NUMBER: 16,773  
; REFERENCE/DOCKET NUMBER: 203424/T7016(C)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627CUSH  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 315 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-129-129-8

Query Match 41.5%; Score 34; DB 1; Length 315;  
Best Local Similarity 37.5%; Pred. NO. 1.6e+02;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 YVPPLLEVGVEKFM 16  
Db 228 FIPLLKADAGVVKKFI 243  
RESULT 26  
US-08-937-972-3  
; Sequence 3, Application US/08937972  
; Patent No. 5932443  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: ANTIGENS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/937,972  
; FILING DATE: Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0400 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 321 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BLADNOT04  
; CLONE: 1318190  
US-08-937-972-3

Query Match 41.5%; Score 34; DB 2; Length 321;  
Best Local Similarity 46.2%; Pred. NO. 1.6e+02;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 VPPLLEVGVEEK 14  
Db 237 VEPILSVGEDE 249

RESULT 27  
US-08-993-260-3  
; Sequence 3, Application US/08993260  
; Patent No. 6031089  
; GENERAL INFORMATION:  
; APPLICANT: Bienkowski, Michael J.  
; TITLE OF INVENTION: NO. 6031089el Sequences of p56, a Proteins Which  
; TITLE OF INVENTION: Affects K-ATP Channels  
; NUMBER OF SEQUENCES: 4

```

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia and Upjohn, Co., Intel. Prop. Law
; ADDRESSEE: (1920-32-LAW)
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: U.S.A.
; ZIP: 49001
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,260
; FILING DATE:
;
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wootton, Thomas A.
; REGISTRATION NUMBER: 35,004
; REFERENCE/DOCKET NUMBER: 6092
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (616) 833-7914
; TELEFAX: (616) 833-8897
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 494 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
;
; US-08-993-260-3
;
;
; Query Match 41.5%; Score 34; DB 3; Length 494;
; Best Local Similarity 54.5%; Pred. No. 2.7e+02;
; Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
;
; QY 6 LLEVGVEEKF 16
; |||||:::
;
; Db 195 LLOVGVTRFI 205
;
;
; RESULT 28
; US-08-946-026-3
; Sequence 3, Application US/08946026
; Patent No. 6034218
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Mitcham, Jennifer L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,026
; FILING DATE: 07-OCT-1997

```

```

;
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
; US-08-946-026-3
;
;
; Query Match 41.5%; Score 34; DB 3; Length 858;
; Best Local Similarity 50.0%; Pred. No. 5.1e+02;
; Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
;
; QY 6 LLEVGVEEKF 15
; |:|:|:|
;
; Db 253 LMSIGLDEKF 262
;
;
; RESULT 29
; US-08-540-406-4
; Sequence 4, Application US/08540406
; Patent No. 5837538
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540,406
; FILING DATE: 06-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-540-406-4
;
;
; Query Match 41.5%; Score 34; DB 2; Length 1311;
; Best Local Similarity 46.7%; Pred. No. 8.4e+02;
; Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 VPPLLLLEVGVEKFM 16  
 ||| | :||: |  
 Db 496 VPFLALGLGVQDMFL 510

## RESULT 30

US-08-656-055-4  
 ; Sequence 4, Application US/08656055  
 ; Patent No. 6027882  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SCOTT, MATHEW P  
 ; APPLICANT: GOODRICH, LISA V  
 ; APPLICANT: JOHNSON, RONALD L  
 ; TITLE OF INVENTION: Patched Genes and their Use  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: IBM PC compatible  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/656,055  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA: 08/540,406  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rowland, Bertram I  
 ; REGISTRATION NUMBER: 20015  
 ; REFERENCE/DOCKET NUMBER: a60190-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-781-1989  
 ; TELEFAX: 415-398-3249  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1311 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-656-055-4

Query Match 41.5%; Score 34; DB 3; Length 1311;  
 Best Local Similarity 46.7%; Pred. No. 8.4e+02;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLLLEVGVEKFM 16  
 ||| | :||: |  
 Db 496 VPFLALGLGVQDMFL 510

## RESULT 31

US-08-954-668-4  
 ; Sequence 4, Application US/08954668  
 ; Patent No. 6172200  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SCOTT, MATHEW P  
 ; APPLICANT: GOODRICH, LISA V  
 ; APPLICANT: JOHNSON, RONALD L  
 ; TITLE OF INVENTION: Patched Genes and their Use  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: Foley, Hoag & Eliot  
 ; STREET: One Post Office Square

CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: ASCII(text)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/954,668  
 ; FILING DATE: 20-Oct-1997  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Vincent, Matthew P.  
 ; REGISTRATION NUMBER: 36709  
 ; REFERENCE/DOCKET NUMBER: SUV-0003.06  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-832-1000  
 ; TELEFAX: 617-832-7000  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1311 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-954-668-4

Query Match 41.5%; Score 34; DB 4; Length 1311;  
 Best Local Similarity 46.7%; Pred. No. 8.4e+02;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLLLEVGVEKFM 16  
 ||| | :||: |  
 Db 496 VPFLALGLGVQDMFL 510

## RESULT 32

PCT-US95-13233-4  
 ; Sequence 4, Application PC/TUS9513233  
 ; GENERAL INFORMATION:  
 ; APPLICANT: THE BOARD OF TRUSTEES OF TEH LELAND STANFORD JUNIOR UNIVERSITY  
 ; TITLE OF INVENTION: Patched Genes and their Use  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/13233  
 ; FILING DATE: 06-OCT-1990  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rowland, Bertram I  
 ; REGISTRATION NUMBER: 20015  
 ; REFERENCE/DOCKET NUMBER: a60190-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-781-1989  
 ; TELEFAX: 415-398-3249  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1311 amino acids  
 ; TYPE: amino acid



```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-13233-4

Query Match      41.5%; Score 34; DB 5; Length 1311;
Best Local Similarity 46.7%; Pred. No. 8.4e+02;
Matches      7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      2 VPPLLLLEVGVEEKFM 16
      || | | : || : | :
Db      496 VPFLALGLGVQDMFL 510

RESULT 33
US-09-413-304-15
; Sequence 15, Application US/09413304
; Patent No. 6207387
; GENERAL INFORMATION:
; APPLICANT: Louis J. Elsas II
; APPLICANT: K. Muralidharan
; TITLE OF INVENTION: MOLECULAR DIAGNOSTICS FOR GALACTOSEMIA
; FILE REFERENCE: 05010.0079
; CURRENT APPLICATION NUMBER: US/09/413,304
; CURRENT FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/No. 6207387e =
US-09-413-304-15

Query Match      40.9%; Score 33.5; DB 4; Length 379;
Best Local Similarity 56.2%; Pred. No. 2.4e+02;
Matches      9; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY      1 YVPPLLLLEVGVEEKFM 16
      | |||| | |||
Db      322 YYPPLLRSATV-RKFM 336

RESULT 34
US-08-905-223-320
; Sequence 320, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duelt, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 320:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 125 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: -18..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 5.6
; OTHER INFORMATION: seq WHFLASFFPRAGC/HG
US-08-905-223-320

Query Match      40.2%; Score 33; DB 4; Length 125;
Best Local Similarity 53.8%; Pred. No. 80;
Matches      7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      3 PPLLEVGVEEKF 15
      | : || | | | | : |
Db      59 PTILLVVGPAEQF 71

RESULT 35
US-08-977-767-1
; Sequence 1, Application US/08977767
; Patent No. 5972684
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Greenwald, Sara
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,767
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0423 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 328 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: OVARNOT03  
CLONE: 2059155  
US-08-977-767-1

Query Match 40.2%; Score 33; DB 2; Length 328;  
Best Local Similarity 46.2%; Pred. No. 2.5e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 YVPPLLLLEVGVEE 13  
:|||||  
Db 83 FLPPRLRLSTGGEK 95

RESULT 36  
US-08-974-546-5  
Sequence 5, Application US/08974546  
Patent No. 5945287  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974.546  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0428  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 340 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1816452  
US-08-974-546-5

Query Match 40.2%; Score 33; DB 2; Length 340;  
Best Local Similarity 53.3%; Pred. No. 2.6e+02;  
Matches 8; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 YVPPLLLLEVGVEEKF 15  
| | | | |

Db 31 YHPDKNKEPGAEKF 45  
RESULT 37  
US-08-987-904A-2  
Sequence 2, Application US/08987904A  
Patent No. 6027917  
GENERAL INFORMATION:  
APPLICANT: Celeste, Anthony J.  
APPLICANT: Murray, Beth  
TITLE OF INVENTION: BONE MORPHOGENETIC PROTEIN (BMP)- 17 AND BMP-18  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: US  
ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/987.904A  
FILING DATE: 10-DEC-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: LAZAR, STEVEN R  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: GI 5307  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8769  
TELEFAX: (617) 876-8581  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 366 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-987-904A-2

Query Match 40.2%; Score 33; DB 3; Length 366;  
Best Local Similarity 46.2%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 4 PLLLEVGVEEKF 16  
|||||  
Db 196 PLLQVSVQREHL 208

RESULT 38  
US-08-987-904A-4  
Sequence 4, Application US/08987904A  
Patent No. 6027917  
GENERAL INFORMATION:  
APPLICANT: Celeste, Anthony J.  
APPLICANT: Murray, Beth  
TITLE OF INVENTION: BONE MORPHOGENETIC PROTEIN (BMP)- 17 AND BMP-18  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: US  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/987,904A  
; FILING DATE: 10-DEC-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LAZAR, STEVEN R  
; REGISTRATION NUMBER: 32,618  
; REFERENCE/DOCKET NUMBER: GI 5307  
; TELEPHONE: (617) 498-8769  
; TELEFAX: (617) 876-8581  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-987-904A-4

Query Match 40.2%; Score 33; DB 3; Length 366;  
Best Local Similarity 46.2%; Pred. No. 2.8e+02;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVVEKFM 16  
||||| | : :  
Db 196 PLLQVSVQREHL 208

RESULT 39  
US-09-080-044-5  
; Sequence 5, Application US/09080044  
; Patent No. 6074649  
; GENERAL INFORMATION:  
; APPLICANT: AUDONNET, Jean-Christophe F.  
; APPLICANT: BAUDU, Philippe G.  
; APPLICANT: RIVIERE, Michel A.  
; TITLE OF INVENTION: RECOMBINANT VACCINE CONTAINING FELINE HERPES VIRUS TYPE  
; TITLE OF INVENTION: 1, PARTICULARLY FOR TREATING FELINE INFECTIOUS  
; TITLE OF INVENTION: PERITONITIS  
; FILE REFERENCE: AUDONNET  
; CURRENT APPLICATION NUMBER: US/09/080,044  
; CURRENT FILING DATE: 1998-05-15  
; EARLIER APPLICATION NUMBER: PCT/FR96/01830  
; EARLIER FILING DATE: 1996-11-19  
; EARLIER APPLICATION NUMBER: 95/14450  
; EARLIER FILING DATE: 1995-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 395  
; TYPE: PRT  
; ORGANISM: Feline herpesvirus 1  
; US-09-080-044-5

Query Match 40.2%; Score 33; DB 3; Length 395;  
Best Local Similarity 54.5%; Pred. No. 3.1e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLLEVGVVEE 13  
|| |||| :  
Db 4 PPSRLEVGIN 14

RESULT 40  
US-09-080-983-5  
; Sequence 5, Application US/09080983  
; Patent No. 6197948  
; GENERAL INFORMATION:

; APPLICANT: Zhu, Hai-Ying  
; APPLICANT: Ling, Kai-Shu  
; APPLICANT: Gonsalves, Dennis  
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS TYPE 2 PROTEINS  
; TITLE OF INVENTION: AND THEIR USES  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/080,983  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/047,194  
; FILING DATE: 20-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1631  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 459 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-080-983-5

Query Match 40.2%; Score 33; DB 4; Length 459;  
Best Local Similarity 37.5%; Pred. No. 3.7e+02;  
Matches 6; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 YVPPLLEVGVVEKFM 16  
| : : | | |  
Db 299 YADDICTDMGFETKFM 314

RESULT 41  
US-08-609-230A-9  
; Sequence 9, Application US/08609230A  
; Patent No. 5866333  
; GENERAL INFORMATION:  
; APPLICANT: Innerarity, Thomas L.  
; APPLICANT: Qian, Xiaobing  
; APPLICANT: Yamanaka, Shinya  
; TITLE OF INVENTION: Screening Methods to Detect mRNA Targets  
; TITLE OF INVENTION: of Editing Enzymes  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA: US/08/609,230A  
;; FILING DATE: 01-MAR-1996  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fitts, Renee A.  
;; REGISTRATION NUMBER: 35,136  
;; REFERENCE/DOCKET NUMBER: 02307U-068100US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 650-326-2400  
;; TELEFAX: 650-326-2422  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 906 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-609-230A-9

Query Match 40.2%; Score 33; DB 2; Length 906;  
Best Local Similarity 63.6%; Pred. No. 8.2e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0;

QY 6 LLEVGVVEKFM 16  
Db 100 LLNVGVESKLI 110

RESULT 42  
US-08-990-140-4  
; Sequence 4, Application US/08990140A  
; Patent No. 6093795

;; GENERAL INFORMATION:  
;; APPLICANT: Olsen, Henrik S.  
;; APPLICANT: Ruben, Steven M.  
;; APPLICANT: Sonenberg, Nahum  
;; APPLICANT: Methot, Nathalie  
;; APPLICANT: Rom, Eran  
;; TITLE OF INVENTION: Human prtl-like Subunit Protein (hPrtl) and Human  
;; FILE REFERENCE: eif4GF-like Protein (p97) Genes  
;; CURRENT APPLICATION NUMBER: US/08/990,140A  
;; CURRENT FILING DATE: 1997-12-12  
;; EARLIER FILING DATE: 1996-12-13  
;; NUMBER OF SEQ ID NOS: 13  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO 4  
;; LENGTH: 907  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-08-990-140-4

Query Match 40.2%; Score 33; DB 3; Length 907;  
Best Local Similarity 63.6%; Pred. No. 8.2e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0;

QY 6 LLEVGVVEKFM 16  
Db 100 LLNVGVESKLI 110

RESULT 43  
US-08-810-712-7  
; Sequence 7, Application US/08810712G  
; Patent No. 6160106

;; GENERAL INFORMATION:  
;; APPLICANT: Yeda Research and Development Co. LTD  
;; TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
;; USE OF SAID GENES AND PROTEINS

;; FILE REFERENCE: sequencelist  
;; CURRENT APPLICATION NUMBER: US/08/810,712G  
;; CURRENT FILING DATE: 1997-03-03  
;; EARLIER APPLICATION NUMBER: PCT/US94/11598  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fitts, Renee A.  
;; REGISTRATION NUMBER: 35,136  
;; REFERENCE/DOCKET NUMBER: 02307U-068100US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 650-326-2400  
;; TELEFAX: 650-326-2422  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 906 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-810-712-7

Query Match 40.2%; Score 33; DB 4; Length 940;  
Best Local Similarity 63.6%; Pred. No. 8.6e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0;

QY 6 LLEVGVVEKFM 16  
Db 133 LLNVGVESKLI 143

RESULT 44  
US-08-460-900C-62  
; Sequence 62, Application US/08460900C  
; Patent No. 6165747

;; GENERAL INFORMATION:  
;; APPLICANT: Ingham, Phillip W.  
;; APPLICANT: McMahon, Andrew P.  
;; APPLICANT: Tabin, Clifford J.  
;; APPLICANT: Bumcrot, David A.  
;; APPLICANT: Marti-Gorostiza, Elisa  
;; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
;; NUMBER OF SEQUENCES: 62  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
;; STREET: One Post Office Square  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02109

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/460,900C  
;; FILING DATE: 5-JUNE-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/435,093  
;; FILING DATE: 4-MAY-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/356,060  
;; FILING DATE: 14-DEC-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/176,427  
;; FILING DATE: 30-DEC-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Vincent, Matthew P.  
;; REGISTRATION NUMBER: 36,709  
;; REFERENCE/DOCKET NUMBER: HMV-006.05  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 832-1000  
;; TELEFAX: (617) 832-7000  
;; INFORMATION FOR SEQ ID NO: 62:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1299 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein



US-08-460-900C-62

Query Match 40.2%; Score 33; DB 4; Length 1299;  
Best Local Similarity 53.3%; Pred. NO. 1.3e+03;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEKF 16  
|||:|:|:|  
Db 494 VPFLALGLGVDFH 508

RESULT 45  
US-08-540-406-10  
; Sequence 10, Application US/08540406  
; Patent No. 5837538  
; GENERAL INFORMATION:  
; APPLICANT: SCOTT, MATHEW P  
; APPLICANT: GOODRICH, LISA V  
; APPLICANT: JOHNSON, RONALD L  
; TITLE OF INVENTION: Patched Genes and their Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/540,406  
; FILING DATE: 06-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20015  
; REFERENCE/DOCKET NUMBER: a60190-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1434 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-540-406-10

Query Match 40.2%; Score 33; DB 2; Length 1434;  
Best Local Similarity 46.7%; Pred. NO. 1.4e+03;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEKF 16  
:|:|:|:|:  
Db 489 LPFLALGVGVDFL 503

Search completed: June 28, 2001, 11:56:09  
Job time: 190 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 42.81 Seconds  
(without alignments)  
28.470 Million cell updates/sec

Title: US-09-439-313-554  
Perfect score: 82  
Sequence: 1 YVPLLLLEVGVEKEFM 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	45	54.9	299	2	PT0060	N-acetylphosphinot
2	45	54.9	299	2	A47031	bialaphos acetylhy
3	43	52.4	545	2	JU0341	intercellular adhe
4	42	51.2	367	2	B72394	citrate synthase -
5	42	51.2	449	2	T17419	probable alkylhali
6	42	51.2	528	2	T31905	hypothetical prote
7	42	51.2	700	2	A83434	conserved hypothet
8	41	50.0	237	2	G65084	hypothetical prote
9	41	50.0	260	1	JS0635	rRNA (adenine-N6-)
10	41	50.0	309	2	G83383	probable esterase/
11	41	50.0	436	2	B70473	protoporphyrinogen
12	41	50.0	3472	2	T31308	hypothetical 367K
13	40	48.8	108	2	S66990	probable membrane
14	40	48.8	171	2	E86820	16S rRNA processin
15	40	48.8	243	2	A69449	heme biosynthesis
16	40	48.8	339	2	S20062	heat shock protein
17	40	48.8	348	2	B72493	hypothetical prote
18	40	48.8	687	2	G81970	probable glycine--
19	40	48.8	687	2	E81027	glycyl-tRNA synth
20	40	48.8	737	2	C70132	hypothetical prote
21	40	48.8	916	2	D83093	secretion protein
22	40	48.8	2114	2	E96505	hypothetical prote
23	39	47.6	162	2	B84152	hypothetical prote
24	39	47.6	297	2	S49885	probable membrane
25	39	47.6	333	2	T48739	probable atp-speci
26	39	47.6	419	2	T29201	hypothetical prote
27	39	47.6	441	2	F86708	citrate (si)-synth
28	39	47.6	531	2	T23835	hypothetical prote
29	39	47.6	570	2	E84825	probable protein k

30	39	47.6	606	2	B81729	glutamine--fructos
31	39	47.6	722	2	A82617	glycyl-tRNA synthe
32	39	47.6	800	2	T26683	hypothetical prote
33	39	47.6	840	2	G85648	probable usher pro
34	39	47.6	1963	2	T49914	callose synthase c
35	38	46.3	227	2	C83046	probable transcrip
36	38	46.3	277	2	T34993	probable oxidoredu
37	38	46.3	387	2	G72288	conserved hypothet
38	38	46.3	419	2	T15088	hypothetical prote
39	38	46.3	449	2	S76839	hypothetical prote
40	38	46.3	490	2	T26171	hypothetical prote
41	38	46.3	506	2	S32561	cysteine proteinas
42	38	46.3	561	2	F75191	hypothetical prote
43	38	46.3	635	2	B72215	hypothetical prote
44	38	46.3	700	2	F64078	translation elonga
45	38	46.3	941	2	T37626	DNA mismatch repai
46	38	46.3	1930	2	F86200	protein F12K11.17
47	38	46.3	2140	2	T18543	probable cell-adhe
48	38	46.3	2561	1	I40456	peptide synthetase
49	37	45.1	115	1	VKLJND	trans-regulatory s
50	37	45.1	139	2	B64010	hypothetical prote
51	37	45.1	164	2	D84152	hypothetical prote
52	37	45.1	179	2	E65134	hypothetical prote
53	37	45.1	179	2	B86004	hypothetical prote
54	37	45.1	207	2	C36961	hypothetical prote
55	37	45.1	225	2	E69350	hypothetical prote
56	37	45.1	266	2	T36341	probable esterase
57	37	45.1	288	2	G84782	hypothetical prote
58	37	45.1	297	2	G81329	probable curved-DN
59	37	45.1	433	1	ITHUC	alpha-1-antichymot
60	37	45.1	437	2	T30897	hypothetical prote
61	37	45.1	446	2	T35627	probable integral
62	37	45.1	450	2	S01367	inner membrane pro
63	37	45.1	559	2	T33168	hypothetical prote
64	37	45.1	596	2	T23193	hypothetical prote
65	37	45.1	621	2	A84933	glutamine--fructos

ALIGNMENTS

RESULT 1  
PT0060  
N-acetylphosphinothricin-tripetide-deacetylase - Streptomyces viridochromogenes  
C;Species: Streptomyces viridochromogenes  
C;Date: 31-Mar-1992 #sequence\_revision 26-Apr-1996 #text\_change 19-May-2000  
C;Accession: S20686; PT0060  
R;Alijah, R.; Hillemann, D.; Nussbaumer, B.; Pelzer, S.; Wohlleben, W.  
submitted to the EMBL Data Library, March 1992  
A;Description: Gene disruption and gene replacement analysis of a 4 kb BamHI fragment  
A;Reference number: S20683  
A;Accession: S20686  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-299 <ALI>  
A;Cross-references: EMBL:X65195; NID:g47997; PIDN:CAA46315.1; PID:g48001  
R;Wohlleben, W.; Arnold, W.; Broer, I.; Hillemann, D.; Strauch, E.; Puehler, A.  
Gene 70, 25-37, 1988  
A;Title: Nucleotide sequence of the phosphinothricin N-acetyltransferase gene from St  
A;Reference number: JT0409; MUID:89196914  
A;Accession: PT0060  
A;Molecule type: DNA  
A;Residues: 1-164 <WOH>  
A;Cross-references: GB:M22827; NID:g295177; PIDN:AAA72710.1; PID:g295180  
A;Experimental source: strain Tue 494  
C;Superfamily: probable lipolytic protein ybaC

Query Match 54.9%; Score 45; DB 2; Length 299;  
Best Local Similarity 66.7%; Pred. No. 2.6;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 VPPLLEVGVEE 13







A:Accession: B70473  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-436 <AQF>  
A:Cross-references: GB:AE000768; NID:g2984249; PIDN:AAC07778.1; PID:g2984251; GB:AE00065  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: hemG

Query Match 50.0%; Score 41; DB 2; Length 436;  
Best Local Similarity 53.3%; Pred. No. 21;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLELVGVVEKF 15  
| | : : | | | | |  
Db 279 YPPVVVVNVGVGEKF 293

RESULT 12  
T31308  
hypothesis protein 367K protein - Cenarchaeum symbiosum  
C:Species: Cenarchaeum symbiosum  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T31308  
R:Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.  
J. Bacteriol. 180, 5003-5009, 1998  
A:Title: Genomic analysis reveals chromosomal variation in natural populations of the un  
A:Reference number: Z20994; MUID:98422450  
A:Accession: T31308  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3472 <SCH>  
A:Cross-references: EMBL:AF083072; NID:g3599393; PID:g3599394; PIDN:AAC62699.1  
C:Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 50.0%; Score 41; DB 2; Length 3472;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLELVGVVEE 13  
| | | | | | | |  
Db 1169 LPPLFLGVGAEE 1180

RESULT 13  
S66990  
probable membrane protein YOR105w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O3217  
C:Species: Saccharomyces cerevisiae  
C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: S66990  
R:Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansoorge, W.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S66965  
A:Accession: S66990  
A:Molecule type: DNA  
A:Residues: 1-108 <VOS>  
A:Cross-references: EMBL:Z75012; NID:g1420285; PID:g1420287; GSPDB:GN00015; MIPS:YOR105w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: MIPS:YOR105w  
A:Map position: 15R  
C:Superfamily: Saccharomyces probable membrane protein YOR105w  
C:Keywords: transmembrane protein  
F:45-61/Domain: transmembrane #status predicted <TMM>

Query Match 48.8%; Score 40; DB 2; Length 108;  
Best Local Similarity 40.0%; Pred. No. 6.5;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLELVGVVEKFM 16  
| | | | | | : : | :  
Db 69 LPPLIIEVHIQPKII 83

RESULT 14  
E86820  
16S rRNA processing protein [imported] - Lactococcus lactis subsp. lactis (strain IL1  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: E86820  
R:Boletini, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh  
Genome Res. in press, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium.  
A:Reference number: A86625  
A:Accession: E86820  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-171 <STO>  
A:Cross-references: GB:AE005176; NID:gl2724568; PIDN:AAK05663.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: rimM

Query Match 48.8%; Score 40; DB 2; Length 171;  
Best Local Similarity 46.2%; Pred. No. 11;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPPLLELVGVVEE 13  
| | | | | | | :  
Db 146 YIPPVILNVVDNQ 158

RESULT 15  
A69449  
heme biosynthesis protein (nirH) homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: A69449  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes,  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
A:Reference number: A69250; MUID:98049343  
A:Accession: A69449  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-243 <KLE>  
A:Cross-references: GB:AE000992; GB:AE000782; NID:g2689315; PIDN:AAB89652.1; PID:g2684

Query Match 48.8%; Score 40; DB 2; Length 243;  
Best Local Similarity 61.5%; Pred. No. 16;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLELVGVVEKFM 16  
| | | | | | | :  
Db 97 PLLRELGFEEFV 109

RESULT 16  
S20062  
heat shock protein dnaJ homolog - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 20-Aug-1999  
C:Accession: S20062; S18086  
R:Raabe, T.; Manley, J.L.  
Nucleic Acids Res. 19, 6645, 1991  
A:Title: A human homologue of the Escherichia coli DnaJ heat- shock protein.

A;Reference number: S20062; MUID:92093635  
A;Accession: S20062  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-339 <RAA1>  
A;Cross-references: EMBL:X62421  
R;Raabe, T.; Manley, J.L.  
Submitted to the EMBL Data Library, October 1991  
A;Description: Primary structure of a human homologue to the Escherichia coli DnaJ prote  
A;Reference number: S18086  
A;Accession: S18086  
A;Molecule type: mRNA  
A;Residues: 1-2, 'K', 4-10, 'Q', 12-33, 'K', 35, 'K', 37-42, 'K', 44, 'K', 46-58, 'K', 60-243, 'N', 245-  
A;Cross-references: EMBL:X62421; NID:g30850; PIDN:CAA44287.1; PID:g30851  
C;Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
F;4-67/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 48.8%; Score 40; DB 2; Length 339;  
Best Local Similarity 56.2%; Pred. No. 23;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKFM 16  
I I I I I I I I I I  
Db 30 YHPDLNLEPGAEELFL 45

RESULT 17  
B72493  
hypothetical protein APE2585 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C;Accession: B72493  
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A;Reference number: A72450; MUID:99310339  
A;Accession: B72493  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-348 <KAW>  
A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81602.1; PID:d1045388; PID:g510  
A;Experimental source: strain K1.  
C;Genetics:  
A;Gene: APE2585

Query Match 48.8%; Score 40; DB 2; Length 348;  
Best Local Similarity 81.8%; Pred. No. 24;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLEVGVEEKF 15  
I I I I I I I I I I  
Db 98 LLLRVGVEEYF 108

RESULT 18  
G81970  
probable glycine--tRNA ligase (EC 6.1.1.14) beta chain NMA0523 [imported] - Neisseria me  
C;Species: Neisseria meningitidis  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C;Accession: G81970  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A;Reference number: A81775; MUID:20222556  
A;Accession: G81970  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-687 <PAR>  
A;Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83815.1; PID:g737926

A;Experimental source: serogroup A, strain Z2491  
C;Genetics:  
A;Gene: glys; NMA0523  
C;Superfamily: glycine--tRNA ligase beta chain  
C;Keywords: ligase

Query Match 48.8%; Score 40; DB 2; Length 687;  
Best Local Similarity 53.8%; Pred. No. 52;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEKFM 16  
I::I I I I I I  
Db 259 PVVLEAGFEEHFL 271

RESULT 19  
E81027  
glycyl-tRNA synthetase, beta chain NMB1930 [imported] - Neisseria meningitidis (strai  
C;Species: Neisseria meningitidis  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C;Accession: E81027  
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;  
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A;Reference number: A81000; MUID:20175755  
A;Accession: E81027  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-687 <TET>  
A;Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AAF42259.1; PID:g722  
A;Experimental source: serogroup B, strain MC58  
C;Genetics:  
A;Gene: NMB1930  
C;Superfamily: glycine--tRNA ligase beta chain

Query Match 48.8%; Score 40; DB 2; Length 687;  
Best Local Similarity 53.8%; Pred. No. 52;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEKFM 16  
I::I I I I I I  
Db 259 PVVLEAGFEEHFL 271

RESULT 20  
C70132  
hypothetical protein BB0259 - Lyme disease spirochete  
C;Species: Borrelia burgdorferi (Lyme disease spirochete)  
C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C;Accession: C70132  
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A;Authors: Smith, H.O.; Venter, J.C.  
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A;Reference number: A70100; MUID:98065943  
A;Accession: C70132  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-737 <KLE>  
A;Cross-references: GB:AE001136; GB:AE000783; NID:g2688152; PIDN:AAC66641.1; PID:g268  
A;Experimental source: strain B31

Query Match 48.8%; Score 40; DB 2; Length 737;  
Best Local Similarity 50.0%; Pred. No. 56;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEKFM 16  
 Db 94 YVAYLFKKIGFEKFKV 109

RESULT 21  
 D83093  
 Secretion protein SecA PA4403 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: D83093  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: D83093  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-916 <STO>  
 A:Cross-references: GB:AE004856; GB:AE004091; NID:g9950633; PIDN:AAG07791.1; GSPDB:GN001  
 A:Experimental source: Strain PA01  
 C:Genetics:  
 A:Gene: secA; PA4403  
 C:Superfamily: preprotein translocase secA

Query Match 48.8%; Score 40; DB 2; Length 916;  
 Best Local Similarity 50.0%; Pred. No. 72;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEKFM 16  
 Db 464 YVSKLLQEAQIEHKVL 479

RESULT 22  
 E96505  
 Hypothetical protein T7023.25 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E96505  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: E96505  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2114 <STO>  
 A:Cross-references: GB:AE005173; NID:g1120821; PIDN:AAG31000.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T7023.25  
 A:Map position: 1

Query Match 48.8%; Score 40; DB 2; Length 2114;  
 Best Local Similarity 60.0%; Pred. No. 1.8e+02;  
 Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 VPPLL--LEVGVEEK 14  
 Db 491 IPPLLQILLETGVSK 505

RESULT 23  
 B84152  
 Hypothetical protein BH4018 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
 C:Accession: B84152  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
 A:Reference number: A83650; MUID:20263314  
 A:Accession: B84152  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-162 <STO>  
 A:Cross-references: GB:AP001520; GB:BA0000004; NID:g10176401; PIDN:BAB07737.1; GSPDB:G  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH4018

Query Match 47.6%; Score 39; DB 2; Length 162;  
 Best Local Similarity 54.5%; Pred. No. 16;  
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEK 14  
 Db 151 PVLIELGLEQK 161

RESULT 24  
 S49885  
 Probable membrane protein YIL124w - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein YI8277.05  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 13-Jan-1995 #sequence\_revision 20-Feb-1995 #text\_change 05-Nov-1999  
 C:Accession: S49885  
 R:Hamlyn, N.; Church, C.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S49881  
 A:Accession: S49885  
 A:Molecule type: DNA  
 A:Residues: 1-297 <HAM>  
 A:Cross-references: GB:Z47047; EMBL:Z46833; NID:g603997; PID:g763222; GSPDB:GN000009;  
 C:Genetics:  
 A:Gene: MIPS:YIL124w  
 A:Map position: 9L  
 A:Superfamily: retinol dehydrogenase; short-chain alcohol dehydrogenase homology  
 C:Keywords: transmembrane protein  
 F:10-188/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 F:255-271/Domain: transmembrane #status predicted <TMM>

Query Match 47.6%; Score 39; DB 2; Length 297;  
 Best Local Similarity 53.3%; Pred. No. 31;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEK 15  
 Db 265 FVPYWLLEKGLSKKF 279

RESULT 25  
 T48739  
 Probable atp-specific succinyl-coa synthetase alpha subunit [imported] - Neurospora c  
 N:Alternate names: protein 8D4.130  
 C:Species: Neurospora crassa  
 C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Sep-2000  
 C:Accession: T48739  
 R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
 submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z24541  
 A:Accession: T48739

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-333 <SCH>  
A;Cross-references: EMBL:AL353819; GSPDB:GN00112; NCSP:8D4.130  
A;Experimental source: cosmid contig 8D4; strain 74  
C;Genetics:  
A;Gene: NCSP:8D4.130  
A;Map position: 2  
A;Introns: 56/2; 66/1; 108/2; 138/1  
C;Superfamily: succinate--CoA ligase (ADP-forming) alpha chain

Query Match 47.6%; Score 39; DB 2; Length 333;  
Best Local Similarity 53.8%; Pred. No. 35;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEE 13  
:||| | |:  
Db 106 FVPPLAAGIEE 118

RESULT 26  
T29201  
hypothetical protein T03F1.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T29201  
R;Du, Z.; Le, T.T.  
submitted to the EMBL Data Library, February 1997  
A;Description: The sequence of C. elegans cosmid T03F1.  
A;Reference number: Z20586  
A;Accession: T29201  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-419 <DUZ>  
A;Cross-references: EMBL:U88169; PIDN:AAB42231.1; GSPDB:GN00019; CESP:T03F1.1  
A;Experimental source: strain Bristol N2; clone T03F1  
C;Genetics:  
A;Gene: CESP:T03F1.1  
A;Map position: 1  
A;Introns: 147/2; 238/3

Query Match 47.6%; Score 39; DB 2; Length 419;  
Best Local Similarity 46.2%; Pred. No. 45;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEK 14  
:|||: |:  
Db 240 VPPLVVGIDER 252

RESULT 27  
F86708  
citrate (si)-synthase (EC 4.1.3.7) [imported] - Lactococcus lactis subsp. lactis (strain  
C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 04-Apr-2001  
C;Accession: F86708  
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich  
Genome Res. in press, 2001  
A;Title: The complete genome sequence of the lactic acid bacterium.  
A;Reference number: A86625  
A;Accession: F86708  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-441 <STO>  
A;Cross-references: GB:AE005176; NID:g12723577; PIDN:AAK04768.1; GSPDB:GN00146  
A;Experimental source: strain IL1403  
C;Genetics:  
A;Gene: gltA  
C;Superfamily: citrate (si)-synthase  
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 47.6%; Score 39; DB 2; Length 441;  
Best Local Similarity 53.3%; Pred. No. 48;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKF 15  
|| | | |:  
Db 335 YVETLVKEKLEEEF 349

RESULT 28  
T23835  
hypothetical protein M88.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C;Accession: T23835  
R;Sulston, J.  
submitted to the EMBL Data Library, June 1994  
A;Reference number: Z19806  
A;Accession: T23835  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-531 <WIL>  
A;Cross-references: EMBL:Z34802; PIDN:CAA84336.1; GSPDB:GN00021; CESP:M88.1  
A;Experimental source: clone M88  
C;Genetics:  
A;Gene: CESP:M88.1  
A;Map position: 3  
A;Introns: 30/3; 71/1; 220/2; 251/3; 288/3; 381/3; 437/2  
C;Superfamily: glucuronosyltransferase

Query Match 47.6%; Score 39; DB 2; Length 531;  
Best Local Similarity 53.3%; Pred. No. 59;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKF 15  
|||::| | |  
Db 187 YVPPLMESDDEMGF 201

RESULT 29

E84825  
probable protein kinase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: E84825  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487  
A;Accession: E84825  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-570 <STO>  
A;Cross-references: GB:AE002093; NID:g6598931; PIDN:AAF18726.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g40120  
A;Map position: 2

Query Match 47.6%; Score 39; DB 2; Length 570;  
Best Local Similarity 54.5%; Pred. No. 64;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEK 14  
| ||::|:  
Db 188 PFLDIGLEDK 198

RESULT 30



B81729  
glutamine--fructose-6-phosphate transaminase (isomerizing) (EC 2.6.1.16) TC0203 [similar  
N:Alternate names: glucosamine fructose-6-phosphate aminotransferase [misnomer]  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MOPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 21-Jul-2000  
C:Accession: B81729  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: B81729  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-606 <TET>  
A:Cross-references: GB:AE002160; NID:g7190237; PIDN:AAF39075.1; PID:g719024  
A:Experimental source: Strain Nigg (MOPn)  
C:Genetics:  
A:Gene: TC0203  
C:Superfamily: glutamine--fructose-6-phosphate aminotransferase (isomerizing)  
C:Keywords: aminotransferase; intramolecular oxidoreductase; isomerase  
F:2-606/Product: glutamine--fructose-6-phosphate transaminase (isomerizing) #status pred  
F:2/Active site: Cys #status predicted

Query Match 47.6%; Score 39; DB 2; Length 606;  
Best Local Similarity 50.0%; Pred. No. 68;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEKFM 16  
I I I I I I I I I I  
Db 175 YESPLILGLGEEVFI 190

RESULT 31  
A82617  
glycyl-tRNA synthetase beta chain XF1959 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: A82617  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: A82617  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-722 <SIM>  
A:Cross-references: GB:AE004015; GB:AE003849; NID:g9107057; PIDN:AAF84761.1; GSPDB:GN001  
A:Experimental source: Strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, B  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshahako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1959  
C:Superfamily: glycine--trNA ligase beta chain

Query Match 47.6%; Score 39; DB 2; Length 722;  
Best Local Similarity 70.0%; Pred. No. 83;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PLLLEVGVEE 13  
I I I I I I I I I I  
Db 6 PLLIELGTEE 15

RESULT 32  
T26683  
hypothetical protein Y38F1A.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T26683  
R:Wallis, J.  
submitted to the EMBL Data Library, October 1998  
A:Reference number: Z20253  
A:Accession: T26683  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-800 <WIL>  
A:Cross-references: EMBL:AL032639; PIDN:CAA21636.1; GSPDB:GN00020; CESP:Y38F1A.3  
A:Experimental source: clone Y38F1A  
C:Genetics:  
A:Gene: CESP:Y38F1A.3  
A:Map position: 2  
A:Introns: 13/2; 100/2; 132/3; 220/3; 257/2; 521/2; 642/3; 710/3

Query Match 47.6%; Score 39; DB 2; Length 800;  
Best Local Similarity 46.7%; Pred. No. 93;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEKFM 16  
I I I I I I I I I I  
Db 232 VPFLILSIGVDDVFI 246

RESULT 33  
G85648  
probable usher protein Z1536 [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: G85648  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85648  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-840 <STO>  
A:Cross-references: GB:AE005174; NID:g12514401; PIDN:AAG55651.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z1536  
C:Superfamily: outer membrane usher protein fimD

Query Match 47.6%; Score 39; DB 2; Length 840;  
Best Local Similarity 50.0%; Pred. No. 98;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEKFM 16  
I I I I I I I I I I  
Db 157 YVPPSELDTGASLAFM 172

RESULT 34  
T49914  
callose synthase catalytic subunit-like protein - Arabidopsis thaliana  
N:Alternate names: protein T24H18.170  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C;Accession: T49914  
R;Bevan, M.; Robben, J.; Grymonprez, B.; Volckaert, G.; Bancroft, I.; Mewes, H.W.; Rudd,  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z25024  
A;Accession: T49914  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1963 <BEV>  
A;Cross-references: EMBL:AL353013; GSPDB:GN000063; ATSP:T24H18.170  
A;Experimental source: cultivar Columbia; BAC clone T24H18  
C;Genetics:  
A;Gene: ATSP:T24H18.170  
A;Map position: 5  
A;Introns: 66/1; 99/3; 141/2; 172/3; 209/3; 248/3; 279/3; 322/2; 364/3; 410/3; 436/2; 46  
1110/2; 1214/1; 1237/1; 1300/2; 1350/3; 1390/3; 1430/2; 1458/2; 1496/1; 1572/3; 1806/1;  
Query Match 47.6%; Score 39; DB 2; Length 1963;  
Best Local Similarity 50.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 4 PILLLEVGVEEKF 15  
:|:|:|:|:|  
Db 1628 PMLMEIGLERGF 1639  
RESULT 35  
C83046  
probable transcription regulator PA4806 [imported] - Pseudomonas aeruginosa (strain PA01  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: C83046  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A;Reference number: A82950; MUID:20437337  
A;Accession: C83046  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-227 <STO>  
A;Cross-references: GB:AE004893; GB:AE004091; NID:g9951063; PIDN:AAG08192.1; GSPDB:GN001  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA4806  
Query Match 46.3%; Score 38; DB 2; Length 227;  
Best Local Similarity 46.7%; Pred. No. 34;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 1 YVPPLLEVGVEEKF 15  
:|:|:|:|:|  
Db 9 HAPTLLEYGIAERF 23  
RESULT 36  
T34993  
probable oxidoreductase - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C;Accession: T34993  
R;Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, December 1998  
A;Reference number: Z21564  
A;Accession: T34993  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-277 <SEE>  
A;Cross-references: EMBL:AL034443; PIDN:CAA22355.1; GSPDB:GN000070; SCOEDB:SC4B5.01c  
A;Experimental source: strain A3(2)  
C;Genetics:

A;Gene: SCOEDB:SC4B5.01c  
C;Superfamily: aldehyde reductase  
Query Match 46.3%; Score 38; DB 2; Length 277;  
Best Local Similarity 63.6%; Pred. No. 43;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 2 VPPLLLEVGVE 12  
|||:| | | |  
Db 5 VPPIILNNGVE 15  
RESULT 37  
G72288  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: G72288  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A;Reference number: A72200; MUID:99287316  
A;Accession: G72288  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-387 <ARN>  
A;Cross-references: GB:AE001772; GB:AE000512; NID:g4981693; PIDN:AAD36223.1; PID:g498  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1147  
C;Superfamily: sensory transduction system regulatory protein; response regulator hom  
Query Match 46.3%; Score 38; DB 2; Length 387;  
Best Local Similarity 50.0%; Pred. No. 63;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 YVPPLLLEVGVE 12  
:|:|:|:|:|  
Db 14 WIKPLLTQIGVE 25  
RESULT 38  
T15088  
hypothetical protein K06A5.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 24-Nov-1999  
C;Accession: T15088  
R;Wamsley, P.  
submitted to the EMBL Data Library, December 1997  
A;Description: The sequence of C. elegans cosmid K06A5.  
A;Reference number: Z18291  
A;Accession: T15088  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-419 <WAM>  
A;Cross-references: EMBL:AF039038; NID:g2736359; PID:g2736364; PIDN:AAB94172.1; GSPDB  
A;Experimental source: strain Bristol N2; clone K06A5  
C;Genetics:  
A;Gene: CESP:K06A5.6  
A;Map position: 1  
A;Introns: 84/3; 265/1; 313/3  
C;Superfamily: acyl-CoA dehydrogenase  
Query Match 46.3%; Score 38; DB 2; Length 419;  
Best Local Similarity 46.2%; Pred. No. 68;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 2 VPPLLLEVGVEEK 14

A:Title: A Plasmodium vinckel cysteine proteinase shares unique features with its Pls  
A:Reference number: S32561; MUID:93250055  
A:Accession: S32561  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-506 <ROS>  
A:Cross-references: GB:L08500  
C:Superfamily: trophozoite cysteine proteinase

Query Match 46.3%; Score 38; DB 2; Length 506;  
Best Local Similarity 43.8%; Pred. No. 84;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 YVPPLLLLEVGVEEKFM 16  
||| : ||| : ||| :  
DB 387 YVGPVTIAVGASEDFV 402

RESULT 42

F75191  
Hypothetical protein PAB0027 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: F75191  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s  
A:Reference number: A75001  
A:Accession: F75191  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-561 <KAW>  
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB48981.1; PID:g545  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0027

Query Match 46.3%; Score 38; DB 2; Length 561;  
Best Local Similarity 43.8%; Pred. No. 95;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLLLEVGVEEKFM 16  
||| : ||| : ||| :  
DB 45 YVAKRLSEIGIEYFL 60

RESULT 43

B72215  
Hypothetical protein TM1762 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72215  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316  
A:Accession: B72215  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-635 <ARN>  
A:Cross-references: GB:AE001814; GB:AE000512; NID:g4982332; PIDN:AAD36826.1; PID:g498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1762  
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology

Query Match 46.3%; Score 38; DB 2; Length 635;  
Best Local Similarity 72.7%; Pred. No. 1.1e+02;

Search completed: June 28, 2001, 11:56:57  
Job time: 238 sec

Query Match	46.3%;	Score 38;	DB 2;	Length 941;
Best Local Similarity	53.8%;	Pred. No. 1.7e+02;		
Matches	7;	Conservative	3;	Mismatches 3; Indels 0; Gaps 0;



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 22.61 Seconds  
(without alignments)  
24.241 Million cell updates/sec

Title: US-09-439-313-554  
Perfect score: 82  
Sequence: 1 YVPPLLEVGVEKEFM 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	54.9	299	BAH_STRHY	Q01109 streptomyce
2	43	52.4	545	ICAI_RAT	Q00238 rattus norv
3	41	50.0	237	YGHS_ECOLI	Q46843 escherichia
4	39	47.6	297	YIM4_YEAST	P40471 saccharomyc
5	39	47.6	722	SYGB_XYLFY	Q9pc26 xylella fas
6	38	46.3	449	Y753_SYNY3	P74635 synecocyst
7	38	46.3	506	CYSP_PLAVN	P46102 plasmodium
8	38	46.3	699	EFG_HAEIN	P43925 haemophilus
9	38	46.3	941	MSH1_SCHPO	O13921 schizosacch
10	38	46.3	2561	PPS1_BACSU	P39845 bacillus su
11	37	45.1	116	REV_HV1ND	P18803 human immun
12	37	45.1	139	Y589_HAEIN	P44020 haemophilus
13	37	45.1	179	YRFC_ECOLI	P45752 escherichia
14	37	45.1	207	YACE_PSEPU	P36644 pseudomonas
15	37	45.1	423	AACT_HUMAN	P01011 homo sapien
16	37	45.1	450	ENVZ_SALTY	P08982 salmonella
17	37	45.1	451	PPOX_MYCLE	Q50008 mycobacteri
18	37	45.1	608	GLMS_BUCAI	P57138 b glucosami
19	37	45.1	679	SYGB_BACSU	P54381 bacillus su
20	37	45.1	1491	AT7A_MOUSE	Q64430 mus musculu
21	37	45.1	1492	AT7A_RAT	P70705 rattus norv
22	37	45.1	2405	DYHA_CHLRE	Q39610 chlamydomon
23	36.5	44.5	483	Y045_MYCGE	P47291 mycoplasma
24	36	43.9	144	Y850_HAEIN	P44060 haemophilus
25	36	43.9	248	YIPA_YEAST	P53039 saccharomyc
26	36	43.9	262	FHUF_ECOLI	P39405 escherichia
27	36	43.9	267	PIV6_ADE40	P48309 human adeno
28	36	43.9	306	KHSE_SYNY3	P73646 synecocyst
29	36	43.9	365	MAP3_SCHPO	P31397 schizosacch
30	36	43.9	368	1 MAP3_SCHPO	O69269 bacillus sp
31	36	43.9	373	CATB_RHOOP	P95608 rhodococcus
32	36	43.9	373	EGON_DROME	P15370 drosophila
33	36	43.9	377	DNAJ_LISMO	Q9s5a3 listeria mo

34	36	43.9	379	1	DNAJ_LACLA	P35514 lactococcus
35	36	43.9	393	1	YSH8_CAEEL	Q09949 caenorhabdi
36	36	43.9	415	1	SACB_ERWAM	Q46554 erwinia amy
37	36	43.9	547	1	SPAK_HUMAN	Q9uew8 homo sapien
38	36	43.9	553	1	SPAK_RAT	O88506 rattus norv
39	36	43.9	556	1	SPAK_MOUSE	Q9z1w9 mus musculu
40	36	43.9	580	1	IUCC_ECOLI	Q47318 escherichia
41	36	43.9	608	1	GLMS_CHLPN	Q9z6u0 c glucosami
42	36	43.9	703	1	EFG_ECOLI	P02996 escherichia
43	36	43.9	732	1	ACPH_HUMAN	P13798 homo sapien
44	36	43.9	930	1	DPO1_HAEIN	P43741 haemophilus
45	36	43.9	985	1	INVA_YERPS	P11922 versinia ps
46	36	43.9	1220	1	PTC1_BRARE	Q98864 brachydanio
47	36	43.9	1226	1	METH_ECOLI	P13009 escherichia
48	36	43.9	3587	1	SRF2_BACSU	Q04747 bacillus su
49	35.5	43.3	502	1	K6B1_HUMAN	P23443 homo sapien
50	35.5	43.3	502	1	K6B1_RAT	P21425 rattus norv
51	35.5	43.3	843	1	MVPA_DICDI	P34118 dictyosteli
52	35.5	43.3	3083	1	POLG_ZYMVR	Q89330 z genome po
53	35	42.7	72	1	RL15_BACLI	P35138 bacillus li
54	35	42.7	100	1	URE3_MYCTU	P50043 mycobacteri
55	35	42.7	130	1	OREX_MOUSE	O55241 mus musculu
56	35	42.7	130	1	OREX_RAT	O55232 rattus norv
57	35	42.7	140	1	FUCU_ECOLI	P11555 escherichia
58	35	42.7	146	1	RL15_BACST	P04452 bacillus st
59	35	42.7	146	1	RL15_BACSU	P19946 bacillus su
60	35	42.7	173	1	DCD_ACIAM	Q02103 acidianus a
61	35	42.7	194	1	CYCY_BRAJA	P30960 bradyrhizob
62	35	42.7	201	1	Y007_NPVAC	P24650 autographa
63	35	42.7	209	1	YDGI_BACSU	P96707 bacillus su
64	35	42.7	253	1	TPIS_BORBU	Q59182 borrelia bu
65	35	42.7	254	1	TPIS_CHLPN	Q9z6j6 chlamydia p

ALIGNMENTS

RESULT 1

BAH_STRHY	ID	BAH_STRHY	STANDARD;	PRT;	299 AA.
AC	Q01109;				
DT	01-APR-1993 (Rel. 25, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	ACETYL-HYDROLASE (EC 3.1.1.-).				
GN	BAH.				
OS	Streptomyces hygroscopicus.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
OX	NCBI_TaxID=1912;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 21705;				
RX	MEDLINE=91294191; PubMed=2066341;				
RA	Raibaud A., Zalacain M., Holt T.G., Tizard R., Thompson C.J.;				
RT	"Nucleotide sequence analysis reveals linked N-acetyl hydrolase,				
RT	thioesterase, transport, and regulatory genes encoded by the				
RT	bialaphos biosynthetic gene cluster of Streptomyces hygroscopicus."				
RL	J. Bacteriol. 173:4454-4463(1991).				
CC	-!- FUNCTION: THIS PROTEIN REMOVES THE N-ACETYL GROUP FROM				
CC	BIALAPHOS AS ONE OF THE FINAL STEPS OF THE BIALAPHOS				
CC	BIOSYNTHETIC PATHWAY.				
CC	-!- PATHWAY: BIALAPHOS BIOSYNTHESIS.				
CC	-!- SIMILARITY: BELONGS TO THE "GDXG" FAMILY OF LIPOLYTIC ENZYMES.				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC					

DR EMBL; M64783; AAA79277.1; -  
DR InterPro; IPR002168; -  
DR PROSITE; PS01173; LIPASE\_GDXG\_HIS; 1.  
DR PROSITE; PS01174; LIPASE\_GDXG\_SER; 1.  
KW Hydrolase; Antibiotic biosynthesis.  
FT ACT\_SITE 73 POTENTIAL.  
FT ACT\_SITE 143 73 POTENTIAL.  
SQ SEQUENCE 299 AA; 32096 MW; 4265C8E6E10FAE97 CRC64;  
  
Query Match 54.9%; Score 45; DB 1; Length 299;  
Best Local Similarity 66.7%; Pred. No. 1.3;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 VPPLLEVGVEE 13  
DB 226 LPPLLIQVSEE 237  
  
RESULT 2  
ICAL\_RAT STANDARD; PRT; 545 AA.  
ID ICAL\_RAT  
AC Q00238;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE INTERCELLULAR ADHESION MOLECULE-1 PRECURSOR.  
GN ICAM1 OR ICAM-1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=1349828;  
RX MEDLINE=92256480; PubMed=1349828;  
RA Kita Y., Takashi T., Iigo Y., Tamatani T., Miyasaka M.;  
RA Horiuchi T.;  
RT "Sequence and expression of rat ICAM-1."  
RL Biochim. Biophys. Acta 1131:108-110(1992).  
CC -!- FUNCTION: ICAM PROTEINS ARE LIGANDS FOR THE LEUKOCYTE ADHESION  
CC -!- LFA-1 PROTEIN.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE ICAM FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D00913; BAA00759.1; -  
DR PIR; S21765; S21765.  
DR PIR; JU0341; JU0341.  
DR HSSP; P05362; 1IC1.  
DR InterPro; IPR003006; -  
DR Pfam; PF00047; ig; 3.  
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Transmembrane;  
Repeat; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 545  
FT DOMAIN 28 492  
FT TRANSMEM 493 517  
FT DOMAIN 518 545  
FT DOMAIN 41 103  
FT DOMAIN 128 193  
FT DOMAIN 230 297  
FT DOMAIN 325 389  
FT DOMAIN 423 476  
FT DISULFID 48 92  
FT DISULFID 52 96  
BY SIMILARITY.  
INTERCELLULAR ADHESION MOLECULE-1.  
EXTRACELLULAR (POTENTIAL).  
POTENTIAL.  
CYTOPLASMIC (POTENTIAL).  
IG-LIKE C2-TYPE DOMAIN.  
IG-LIKE C2-TYPE DOMAIN.  
IG-LIKE C2-TYPE DOMAIN.  
IG-LIKE C2-TYPE DOMAIN.  
IG-LIKE C2-TYPE DOMAIN.  
BY SIMILARITY.

FT DISULFID 135 186 BY SIMILARITY.  
FT DISULFID 237 290 BY SIMILARITY.  
FT DISULFID 332 382 BY SIMILARITY.  
FT DISULFID 430 469 BY SIMILARITY.  
FT SITE 177 179 CELL ATTACHMENT SITE (POTENTIAL).  
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 545 AA; 60141 MW; 30F4546FA4D0CFF4 CRC64;  
  
Query Match 52.4%; Score 43; DB 1; Length 545;  
Best Local Similarity 61.5%; Pred. No. 5.5;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
QY 4 PLLLEVGVEEKFM 16  
DB 224 PDLLEVGTTQKFL 236  
  
RESULT 3  
YGHS\_ECOLI STANDARD; PRT; 237 AA.  
ID YGHS\_ECOLI  
AC Q46843;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 26.3 KDA ATP-BINDING PROTEIN IN GLCC-PITB INTERGENIC  
DE REGION.  
GN YGHS.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
CC -!- SIMILARITY: TO E.COLI YGHR AND YGHT.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U28377; AAA69152.1; -  
DR EMBL; AE000381; AAC76021.1; -  
DR EcoGene; EG13003; YGHS.  
KW Hypothetical protein; ATP-binding.  
FT NP\_BIND 21 28 ATP (POTENTIAL).  
FT SEQUENCE 237 AA; 26346 MW; 69D8AE6673D7DA6F CRC64;  
  
Query Match 50.0%; Score 41; DB 1; Length 237;  
Best Local Similarity 42.9%; Pred. No. 5.2;  
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 YVPPLLEVGVEEK 14
   1:1 ||: ||: ||:
Db 168 YLPVLLIRLGIDEQ 181

RESULT 4
YIM4_YEAST
ID YIM4_YEAST STANDARD; PRT; 297 AA.
AC P40471;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN KGD1-SIM1 INTERGENIC REGION
DE (EC 1.-.-.-).
GN YILL124W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY. STRONG, TO S.POMBE SPAC23D3.11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46833; CAA86868.1; -.
CC SGD; S0001386; YILL124W.
CC InterPro; IPR002198; -.
CC Pfam; PF00106; adh_short; 1.
CC PROSITE; PS00061; ADH_SHORT; 1.
CC KW Hypothetical protein; Oxidoreductase.
CC FT NP_BIND 13 37 NAD OR NADP (BY SIMILARITY).
CC ACT_SITE 157 157 BY SIMILARITY.
CC SEQUENCE 297 AA; 32814 MW; B614C0E0B1FB0CE0 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 297;
Best Local Similarity 53.3%; Pred. No. 15;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEK 15
   || || || ||: ||
Db 265 FVPYWLLEKGLSKKF 279

RESULT 5
SYGB_XYLFA
ID SYGB_XYLFA STANDARD; PRT; 722 AA.
AC Q9PC26;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLYCYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.14) (GLYCINE--TRNA LIGASE
DE BETA CHAIN) (GLYRS).
GN GLYS OR XF1959.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
```

```
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-GLYCINE + TRNA(GLY) = AMP +
CC PYROPHOSPHATE + L-GLYCYL-TRNA(GLY).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004015; AAF84761.1; -.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
CC SEQUENCE 722 AA; 80149 MW; AB6D0E4712B889EE CRC64;

Query Match 47.6%; Score 39; DB 1; Length 722;
Best Local Similarity 70.0%; Pred. No. 38;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PLLLEVGVEE 13
   |||: ||: ||
Db 6 PLLLELGTTEE 15

RESULT 6
Y753_SYNY3
ID Y753_SYNY3 STANDARD; PRT; 449 AA.
AC P74635;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 48.0 kDa protein SLR0753.
GN SLR0753.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
```



```

NCBI_TaxID-1148;
[1]
SEQUENCE FROM N.A.
MEDLINE=97061201; PubMed=8905231;
Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
Miyajima N., Hirose M., Sugiura M., Sasamoto S., Kimura T.,
Hosouchi T., Matsuno A., Murauchi A., Nakazaki N., Naruo K.,
Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
Yamada M., Yasuda M., Tabata S.;
"Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
DNA Res. 3:109-136(1996).
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
-!- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.
P SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; D90917; BAA18751.1; --
Hypothetical protein; Transmembrane; Transport.
TRANSMEM 1 21 POTENTIAL.
TRANSMEM 26 46 POTENTIAL.
TRANSMEM 51 71 POTENTIAL.
TRANSMEM 97 117 POTENTIAL.
TRANSMEM 137 157 POTENTIAL.
TRANSMEM 178 198 POTENTIAL.
TRANSMEM 223 243 POTENTIAL.
TRANSMEM 244 264 POTENTIAL.
TRANSMEM 285 305 POTENTIAL.
TRANSMEM 310 330 POTENTIAL.
TRANSMEM 340 360 POTENTIAL.
TRANSMEM 377 397 POTENTIAL.
TRANSMEM 425 445 POTENTIAL.
SEQUENCE 449 AA; 48003 MW; 9673EBCFE72F483F CRC64;
Query Match 46.3%; Score 38; DB 1; Length 449;
Best Local Similarity 54.5%; Pred. No. 35;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 2 VPPLLEVGVE 12
:|||||:
Db 126 IPPLAQEIGVD 136
-----
RESULT 7
CYP_PLVN STANDARD; PRT; 506 AA.
AC P46102;
Dt 01-NOV-1995 (Rel. 32, Created)
Dt 01-NOV-1995 (Rel. 32, Last sequence update)
Dt 01-NOV-1995 (Rel. 32, Last annotation update)
DE CYSTEINE PROTEINASE PRECURSOR (EC 3.4.22.-).
OS Plasmodium vinckei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5860;
[1]
SEQUENCE FROM N.A.
MEDLINE=93250055; PubMed=8485161;
Rosenthal P.J.;
"A plasmodium vinckei cysteine proteinase shares unique features with
its plasmodium falciparum analogue.";
Biochim. Biophys. Acta 1173:91-93(1993).
-!- FUNCTION: PROBABLY DEGRADES ERYTHROCYTE HEMOGLOBIN.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
PAPAIN FAMILY OF THIOLE PROTEASES.

```

```

CC -1- SIMILARITY: STRONGEST SIMILARITY TO CATHEPSIN L.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L08500; -; NOT_ANNOTATED_CDS.
CC HSSP; P43235; 1ATK.
CC MEROPS; C01.077; -
CC InterPro; IPR000169; -
CC InterPro; IPR000668; -
CC Pfam; PF00112; Peptidase_C1; 3.
CC PRINTS; PR00705; PAPA1N.
CC PROSITE; PS00139; THIO_L_PROTEASE_CYS; 1.
CC PROSITE; PS00639; THIO_L_PROTEASE_HIS; 1.
CC PROSITE; PS00640; THIO_L_PROTEASE_ASN; 1.
CC Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal.
KW SIGNAL
FT ?
FT PROPEP ? 262 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 263 506 CYSTEINE PROTEINASE.
FT ACT_SITE 287 287 BY SIMILARITY.
FT ACT_SITE 419 419 BY SIMILARITY.
FT ACT_SITE 470 470 BY SIMILARITY.
FT DISULFID 284 326 BY SIMILARITY.
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 506 AA; 58255 MW; 7478494461A617F1 CRC64;
CC
CC Query Match 46.3%; Score 38; DB 1; Length 506;
CC Best local Similarity 43.8%; Pred. NO. 40;
CC Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
CC
CC QY 1 YVPPLLLLEGVGVEKFM 16
CC || | : | | | |
CC Db 387 YVGPVTIAGASEDFV 402
CC
CC RESULT 8
CC EFG_HAEIN STANDARD; PRT; 699 AA.
CC AC P43925;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
CC DE ELONGATION FACTOR G (EF-G).
CC GN FUSA OR FUS OR HI0579.
CC OS Haemophilus influenzae.
CC OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC OC Haemophilus
CC OX NCBI_TaxID=727;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=RD / KW20 / ATCC 51907;
CC MEDLINE=95350630; PubMed=7542800;
CC Klebschmann R.D.; Adams M.D.; White O.; Clayton R.A.; Kirkness E.F.;
CC Kierlavage A.R.; Bult C.J.; Tomb J.-F.; Dougherty B.A.; Merrick J.M.;
CC McKenney K.; Sutton G.; Fitzhugh W.; Fields C.A.; Gocayne J.D.;
CC Scott J.D.; Shirley R.; Liu L.-I.; Glodek A.; Kelley J.M.;
CC Weidman J.F.; Phillips C.A.; Spriggs T.; Hedblom E.; Cotton M.D.;
CC Utterback T.R.; Hanna M.C.; Nguyen D.T.; Saudek D.M.; Brandon R.C.;
CC Fine L.D.; Fritchman J.L.; Fuhrmann J.L.; Geoghagen N.S.M.;
CC Gnehm C.L.; McDonald L.A.; Small K.V.; Fraser C.M.; Smith H.O.;
CC Venter J.C.;
CC "Whole-genome random sequencing and assembly of Haemophilus
CC influenzae Rd.;"
CC Science 269:496-512(1995).

```

CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION  
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE  
CC RIBOSOME.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
CC EF-G/EF-2 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U32739; AAC22237.1; -.  
DR HSSP; P13551; IELO.  
DR TIGR; HI0579; -.  
DR InterPro; IPR000640; -.  
DR InterPro; IPR000795; -.  
DR Pfam; PF00679; EFG\_C; 1.  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR PRINTS; PR00315; ELONGATNFCT.  
DR PROSITE; PS00301; EFACITOR\_GTP; 1.  
KW Elongation factor; Protein biosynthesis; GTP-binding.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT NP\_BIND 16 23 GTP (BY SIMILARITY).  
FT NP\_BIND 87 91 GTP (BY SIMILARITY).  
FT NP\_BIND 141 144 GTP (BY SIMILARITY).  
SQ SEQUENCE 699 AA; 77132 MW; FBBAD39C0F62801 CRC64;  
  
Query Match 46.3%; Score 38; DB 1; Length 699;  
Best Local Similarity 66.7%; Pred. No. 56;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 4 PLLLEVGVEKF 15  
Db 167 PLQLPVGAENF 178  
||| |||||  
  
RESULT 9  
MSH1\_SCHPO  
ID MSH1\_SCHPO STANDARD; PRT; 941 AA.  
AC O13921; O13700;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE MUTS PROTEIN HOMOLOG 1.  
GN SPAC23C11.18C OR SPAC13F5.01C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: INVOLVED IN MITOCHONDRIAL DNA REPAIR (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z98559; CAB11169.1; -.  
DR EMBL; Z99091; CAB11774.1; -.

DR InterPro; IPR000432; -.  
DR InterPro; IPR002863; -.  
DR Pfam; PF00488; Muts\_C; 1.  
DR Pfam; PF01624; Muts\_N; 1.  
DR PROSITE; PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
KW DNA repair; ATP-binding; DNA-binding; Mitochondrion.  
FT NP\_BIND 747 754 ATP (POTENTIAL).  
SQ SEQUENCE 941 AA; 106932 MW; 7A1D8F477E1140AB CRC64;  
  
Query Match 46.3%; Score 38; DB 1; Length 941;  
Best Local Similarity 53.8%; Pred. No. 76;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 VPPLLEVGVEEK 14  
Db 60 LPPLLEKVSFQOK 72  
:||||| :|  
  
RESULT 10  
PPS1\_BACSU STANDARD; PRT; 2561 AA.  
AC P39845;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PEPTIDE SYNTHETASE 1.  
GN PPSA OR PPS1.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95227362; PubMed=7711903;  
RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,  
RA Grandi G.;  
RT "A putative new peptide synthase operon in Bacillus subtilis: partial  
RT characterization";  
RL Microbiology 141:645-648(1995).  
CC -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE  
CC (POTENTIAL).  
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z34883; CAA84360.1; -.  
DR EMBL; Z99113; CAB13717.1; -.  
DR HSSP; P14687; 1AMU.  
DR Subtilist; BG10970; ppsA.  
DR InterPro; IPR000255; -.  
DR InterPro; IPR000873; -.  
DR InterPro; IPR001242; -.  
DR Pfam; PF00501; AMP-binding; 2.  
DR Pfam; PF00668; DUF4; 3.  
DR Pfam; PF00550; pp-binding; 2.  
DR PRINTS; PR00154; AMPBINDING.  
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.  
DR PROSITE; PS00455; AMP\_BINDING; 2.  
DR PROSITE; PS50075; ACP\_DOMAIN; 2.  
KW Multifunctional enzyme; Ligase; Repeat; Phosphopantetheine.  
FT DOMAIN 966 1033 ACYL CARRIER (ACP).  
FT DOMAIN 2012 2078 ACYL CARRIER (ACP).  
FT BINDING 2042 2042 PHOSPHOPANTHETHEINE (POTENTIAL).  
SQ SEQUENCE 2561 AA; 289180 MW; 5476CBE4DD882FD2 CRC64;

```

Query Match      46.3%; Score 38; DB 1; Length 2561;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 PLELVGVVEKF 15
Db 2394 PILDMGIPFP 2405

RESULT 11
REV_HVIND STANDARD; PRT; 116 AA.
AC P18803;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE REV PROTEIN (ANTI-REPRESSION TRANSACTIVATOR PROTEIN) (ART/TRS).
GN REV.
OS Human immunodeficiency virus type 1 (NDK isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11695;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90034200; PubMed=2806917;
RA Spire B., Sire J., Zachar V., Rey F., Barre-Sinoussi F., Galibert F.,
RA Hampe A., Chermann J.C.;
RA "Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the
RT human immunodeficiency virus.";
RL Gene 81:275-284(1989).
CC -!- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; ACCUMULATES IN THE NUCLEOLI.
CC -!- PTM: PHOSPHOPROTEIN WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
CC BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
CC -!- MISCELLANEOUS: NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH
CC AIDS, AND IS A HIGHLY CYTOPATHOGENIC STRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27323; AAA44867.1;
CC PIR; JQ0072; VKLJND.
CC HIV; M27323; REV$NDK.
CC InterPro; IPR000625;
CC Pfam; PF00424; REV; 1.
CC Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
SQ SEQUENCE 116 AA; 12983 MW; 40CC1B562B655AEE CRC64;

Query Match      45.1%; Score 37; DB 1; Length 116;
Best Local Similarity 63.6%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLELVGVEE 13
Db 106 PPAVLESGTEE 116

RESULT 12
Y589_HAEIN STANDARD; PRT; 139 AA.
AC P44020;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL PROTEIN HI0589.

Query Match      45.1%; Score 37; DB 1; Length 139;
Best Local Similarity 42.1%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 2; Indels 4; Gaps 1;

QY 2 VPPL-----LLEVGVEEKFM 16
Db 52 ITPLRTDQMVEIGLEKSM 70

RESULT 13
YRFC_ECOLI STANDARD; PRT; 179 AA.
AC P45752;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 20.8 KDA PROTEIN IN HOFQ-MRCA INTERGENIC REGION (F179).
GN YRFC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).

GN HI0589.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: TO E-COLI RSEC AND H-INFLUENZAE HI850.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32740; AAC22246.1;
CC TIGR; HI0589;
CC Hypothetical protein; Transmembrane; Inner membrane.
KW TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
SQ SEQUENCE 139 AA; 15635 MW; 6C26821FAF4DF32D CRC64;

Query Match      45.1%; Score 37; DB 1; Length 139;
Best Local Similarity 42.1%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 2; Indels 4; Gaps 1;

QY 2 VPPL-----LLEVGVEEKFM 16
Db 52 ITPLRTDQMVEIGLEKSM 70
```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U18997; AAA58191.1; -.  
DR EMBL; AF000414; AAC76419.1; -.  
DR EcoGene; EGI2924; yrfC.  
KW Hypothetical protein.  
SQ SEQUENCE 179 AA; 20796 MW; 2A7D86A4B36B10C8 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 179;  
Best Local Similarity 63.6%; Pred. No. 20;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVPPLLEVG 11  
: | ||| | |  
Db 26 FVAPLLAVGI 36

RESULT 14  
YACE\_PSEPU STANDARD; PRT; 207 AA.  
AC P36644;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 23.0 KDA PROTEIN IN XCPA/PILD 3 REGION.  
OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WCS358;  
RX MEDLINE=94131942; PubMed=7905475;  
RA de Groot A., Heijnen I., de Cock H., Filloux A., Tommassen J.;  
RT "Characterization of type IV pilus genes in plant growth-promoting Pseudomonas putida WCS358.";  
RL J. Bacteriol. 176:642-650(1994).  
CC -!- SIMILARITY: BELONGS TO THE UPF0038 FAMILY.  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X74276; CAA52335.1; -.  
DR PIR; S35954; S35954.  
DR InterPro; IPR001977; -.  
DR Pfam; PF01121; UPF0038; 1.  
DR PROSITE; PS01294; UPF0038; 1.  
KW Hypothetical protein; ATP-binding.  
FT NP\_BIND 15 22 ATP (POTENTIAL).  
SQ SEQUENCE 207 AA; 22975 MW; D62B74957FFED098 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 207;  
Best Local Similarity 57.1%; Pred. No. 24;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEK 14  
| | | | | |  
Db 116 YVSPLLIESGQYRK 129

RESULT 15  
AACT\_HUMAN STANDARD; PRT; 423 AA.  
ID AACT\_HUMAN  
AC P01011;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ALPHA-1-ANTICHYMOTRYPSIN PRECURSOR (ACT).  
GN SERPINA3 OR AACT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84080367; PubMed=6606438;  
RA Chandra T., Stackhouse R., Kidd V.J., Robson K.J.H., Woo S.L.C.;  
RT "Sequence homology between human alpha 1-antichymotrypsin, alpha 1-antitrypsin, and antithrombin III.";  
RL Biochemistry 22:5055-5061(1983).  
RN [2]  
RP SEQUENCE OF 87-129 FROM N.A.  
RX MEDLINE=90110106; PubMed=2404007;  
RA Rubin H., Wang Z., Nickbarg E.B., McLarney S., Naidoo N., Schoenberger O.L., Johnson J.L., Cooperman B.S.;  
RT "Cloning, expression, purification, and biological activity of recombinant native and variant human alpha 1-antichymotrypsins.";  
RL J. Biol. Chem. 265:1199-1207(1990).  
RN [3]  
RP SEQUENCE OF 22-423 FROM N.A.  
RA Rubin H.;  
RL Submitted (OCT-1989) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF N-TERMINUS.  
RX MEDLINE=89323223; PubMed=2787670;  
RA Lindmark B., Hilja H., Alan R., Eriksson S.;  
RT "The microheterogeneity of desialylated alpha 1-antichymotrypsin: the occurrence of two amino-terminal isoforms, one lacking a His-Pro dipeptide.";  
RL Biochim. Biophys. Acta 997:90-95(1989).  
RN [5]  
RP ACTIVE SITE.  
RX MEDLINE=84032476; PubMed=6556193;  
RA Morii M., Travis J.;  
RT "Amino acid sequence at the reactive site of human alpha 1-antichymotrypsin.";  
RL J. Biol. Chem. 258:12749-12752(1983).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE=91202538; PubMed=2016749;  
RA Baumann U., Huber R., Bode W., Grosse D., Lesjak M., Laurell C.-B.;  
RT "Crystal structure of cleaved human alpha 1-antichymotrypsin at 2.7-A resolution and its comparison with other serpins.";  
RL J. Mol. Biol. 218:595-606(1991).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).  
RX MEDLINE=96433079; PubMed=8836107;  
RA Lukacs C.M., Zhong J.Q., Plotnick M.I., Rubin H., Cooperman B.S., Christianson D.W.;  
RT "Arginine substitutions in the hinge region of antichymotrypsin affect serpin beta-sheet rearrangement.";  
RL Nat. Struct. Biol. 3:888-893(1996).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
RX MEDLINE=98198038; PubMed=9521649;  
RA Lukacs C.M., Rubin H., Christianson D.W.;  
RT "Engineering an anion-binding cavity in antichymotrypsin modulates the 'spring-loaded' serpin-protease interaction.";  
RL Biochemistry 37:3297-3304(1998).  
RN [9]  
RP VARIANTS.  
RX MEDLINE=94063919; PubMed=8244391;



RA Poller W., Faber J.-P., Weidinger S., Tief K., Scholz S., Fischer M.,  
RA Olek K., Kirchgesser M., Heidtmann H.-H.,  
RT "A leucine-to-proline substitution causes a defective alpha 1-  
RT antichymotrypsin allele associated with familial obstructive lung  
RT disease.";  
RL Genomics 17:740-743(1993).  
RN [10]  
RP VARIANT VAL-401.  
RX MEDLINE=92316200; PubMed=1618300;  
RA Tsuda M., Sei Y., Yamamura M., Yamamoto M., Shinohara Y.,  
RT "Detection of a new mutant alpha-1-antichymotrypsin in patients with  
RT occlusive-cerebrovascular disease.";  
RL FEBS Lett. 304:66-68(1992).  
CC -1- FUNCTION: ALTHOUGH ITS PHYSIOLOGICAL FUNCTION IS UNCLEAR, IT CAN  
CC INHIBIT NEUTROPHIL CATHEPSIN G AND MAST CELL CHYMASE, BOTH OF  
CC WHICH CAN CONVERT ANGIOTENSIN I TO THE ACTIVE ANGIOTENSIN II.  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -1- TISSUE SPECIFICITY: PLASMA. SYNTHESIZED IN THE LIVER. LIKE  
CC THE RELATED ALPHA-1-ANTITRYPSIN, ITS CONCENTRATION INCREASES IN  
CC THE ACUTE PHASE OF INFLAMMATION OR INFECTION.  
CC -1- DISEASE DEFICIENCY IN ACT CAN BE A CAUSE OF CHRONIC OBSTRUCTIVE  
CC PULMONARY DISEASE (COPD) OR OF OCCLUSIVE CEREBROVASCULAR DISEASE.  
CC -1- MISCELLANEOUS: ALPHA-1-ANTICHYMOTRYPSIN CAN BIND DNA.  
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-4 IS THE INITIATOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: K01500; AAA51543.1; --  
CC EMBL: J05176; AAA51560.1; --  
CC PIR: A01246; ITHUC.  
CC PIR: A34934; A34934.  
CC PIR: S14806; S14806.  
CC PDB: 2ACH; 15-JUL-93.  
CC PDB: 3CAA; 25-FEB-98.  
CC PDB: 4CAA; 25-FEB-98.  
CC PDB: 1AS4; 25-FEB-98.  
CC GlycoSuiteDB: P01011; --  
CC SWISS-2DPAGE; P01011; HUMAN.  
CC MIM: 107280; --  
CC InterPro: IPR000215; --  
CC Pfam: PF00079; serpin; 1.  
CC PROSITE: PS00284; SERPIN; 1.  
KW Serpin; Serine protease inhibitor; Plasma; Acute phase; Glycoprotein;  
KW Signal; 3D-structure; Polymorphism; Disease mutation.  
FT SIGNAL 1 23  
FT CHAIN 24 423  
FT ACT\_SITE 383 384  
FT CARBOHYD 93 93  
FT CARBOHYD 127 127  
FT DNA\_BIND 235 237  
FT VARIANT 9 9  
FT  
FT VARIANT 78 78  
FT  
FT VARIANT 167 167  
FT  
FT VARIANT 252 252  
FT  
FT VARIANT 401 401  
FT  
FT CONFLICT 69 69  
FT CONFLICT 102 116  
FT  
FT CONFLICT 123 128  
FT CONFLICT 199 199  
FT CONFLICT 361 363

FT CONFLICT 421 423  
FT HELIX 48 67  
FT TURN 69 70  
FT STRAND 73 75  
FT HELIX 77 89  
FT TURN 90 90  
FT HELIX 93 102  
FT TURN 103 104  
FT TURN 107 109  
FT HELIX 112 126  
FT STRAND 133 144  
FT TURN 145 146  
FT HELIX 151 161  
FT STRAND 164 168  
FT TURN 170 171  
FT HELIX 173 187  
FT TURN 188 189  
FT TURN 201 202  
FT STRAND 203 217  
FT HELIX 223 225  
FT STRAND 227 230  
FT TURN 240 256  
FT TURN 257 260  
FT STRAND 261 268  
FT STRAND 272 279  
FT TURN 281 283  
FT HELIX 284 289  
FT TURN 290 290  
FT STRAND 293 302  
FT STRAND 304 314  
FT STRAND 316 320  
FT STRAND 323 323  
FT HELIX 325 330  
FT TURN 331 332  
FT HELIX 335 337  
FT TURN 339 340  
FT TURN 344 345  
FT TURN 350 351  
FT STRAND 352 365  
FT STRAND 369 382  
FT STRAND 392 394  
FT STRAND 399 405  
FT STRAND 414 417  
FT TURN 421 422  
SQ SEQUENCE 423 AA; 47650 MW; B002F946C86A8951 CRC64;  
  
Query Match 45.1%; Score 37; DB 1; Length 423;  
Best Local Similarity 54.5%; Pred. No. 50;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
  
QY 5 LILEVGVVEKF 15  
DB 327 ILLQLGIEEAF 337  
  
RESULT 16  
ENVZ\_SALTY STANDARD; PRT; 450 AA.  
AC P08982;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE OSMOLARITY SENSOR PROTEIN ENVZ (EC 2.7.3.-).  
GN ENVZ.  
OS Salmonella typhimurium.  
OC Bacteria; proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2;  
RX MEDLINE=89011922; PubMed=2845093;

```
RA Lijestrom P., Laamanen I., Palva E.T.;
RT "Structure and expression of the ompB operon, the regulatory locus
RT for the outer membrane porin regulon in Salmonella typhimurium
RT LT-2.";
RL J. Mol. Biol. 201:663-673(1988).
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM ENVZ/OMPR
CC INVOLVED IN THE REGULATION OF OSMOREGULATION (GENES OMPF & OMPG).
CC ENVZ FUNCTIONS AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT
CC PHOSPHORYLATES OMPR IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X12374; CAA30935.1; -.
CC PIR; S01367; S01367.
CC StyGene; SG10091; envZ.
CC InterPro; IPR000410; -.
CC InterPro; IPR000658; -.
CC Pfam; PF00672; DUF5; 1.
CC Pfam; PF00512; signal; 1.
CC Sensory transduction; Transferase; Kinase; Phosphorylation;
KW Transmembrane; Inner membrane.
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 16 35 PROBABLE.
FT DOMAIN 36 161 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 162 182 PROBABLE.
FT DOMAIN 183 450 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 220 450 TRANSMITTER DOMAIN (POTENTIAL).
FT MOD_RES 243 243 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 450 AA; 50331 MW; 1E424AE0CF78225B CRC64;

Query Match 45.1%; Score 37; DB 1; Length 450;
Best Local Similarity 58.3%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEE 13
:| | | | |
Db 201 IPPPLLEYGASE 212

RESULT 17
PPOX_MYCLE
ID PPOX_MYCLE STANDARD; PRT; 451 AA.
AC Q50008;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PROTOPORPHYRINOGEN OXIDASE (EC 1.3.3.4) (PPO).
GN HEMY.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE 6-ELECTRON OXIDATION OF PROTOPORPHYRINOGEN
CC IX TO FORM PROTOPORPHYRIN IX.
CC -!- CATALYTIC ACTIVITY: PROTOPORPHYRINOGEN-IX + O(2) -> PROTOPORPHYRIN-
CC IX + H(2)O(2).
CC -!- COFACTOR: CONTAINS ONE FAD PER HOMODIMER (BY SIMILARITY).
CC -!- PATHWAY: PENULTIMATE STEP IN HEME AND PORPHYRIN BIOSYNTHESIS.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
```

```
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PROTOPORPHYRINOGEN OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U15181; AAA62958.1; -.
CC Porphyrin biosynthesis; Oxidoreductase; Flavoprotein; FAD;
KW Heme biosynthesis.
FT NP_BIND 10 15 FAD (POTENTIAL).
SQ SEQUENCE 451 AA; 46580 MW; DF76EE1655CA2056 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 451;
Best Local Similarity 53.8%; Pred. No. 53;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEK 14
| | | | |
Db 69 VPALLAELGLSER 81

RESULT 18
GLMS_BUCAI
ID GLMS_BUCAI STANDARD; PRT; 608 AA.
AC P57138;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING]
DE (EC 2.6.1.16) (HEXOSEPHOSPHATE AMINOTRANSFERASE) (D-FRUCTOSE-6-
DE PHOSPHATE AMIDOTRANSFERASE) (GFAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE
DE AMIDOTRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).
GN GLMS OR BU026.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -!- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM,
CC CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A
CC NITROGEN SOURCE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: L-GLUTAMINE + D-FRUCTOSE 6-PHOSPHATE ->
CC L-GLUTAMATE + D-GLUCOSAMINE 6-PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2
CC GATASE DOMAIN.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.
CC GLMS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001118; BAB12753.1; ALT_INIT.
CC PROSITE; PS00443; GATASE_TYPE_II; 1.
KW Transferase; Aminotransferase; Glutamine amidotransferase.
```

```
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
FT ACT_SITE 603 603 ISOMERISATION FRU-6P (BY SIMILARITY).
FT DOMAIN 1 182 GLUTAMINE AMIDOTRANSFERASE.
SQ SEQUENCE 608 AA; 68550 MW; 77A45F96CBD190BC CRC64;

Query Match 45.1%; Score 37; DB 1; Length 608;
Best Local Similarity 46.2%; Pred. No. 72;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEK 16
Db 178 PLIIGLGTENFI 190

RESULT 19
SYGB_BACSU STANDARD; PRT; 679 AA.
ID SYGB_BACSU
AC P54381;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE GLYCYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.14) (GLYCINE--TRNA LIGASE
DE BETA CHAIN) (GLYRS).
GN GLYS
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.; to the EMBL/GenBank/DBJ databases.
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RL -!- CATALYTIC ACTIVITY: ATP + L-GLYCYL-TRNA (GLY) = AMP +
CC PYROPHOSPHATE + L-GLYCYL-TRNA (GLY).
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84432; BAA12485.1; -
CC EMBL; Z99116; CAB14455.1; -
CC EMBL; Z99117; CAB14468.1; -
CC Subtilist; BG11658; glyS.
CC InterPro; IPR002311; -
CC Pfam; PF02092; trna_synth_2f; 1.
CC PRINTS; PR01045; TRNASYNTHGB.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ SEQUENCE 679 AA; 76232 MW; ACC96723A85D92BB CRC64;

Query Match 45.1%; Score 37; DB 1; Length 679;
Best Local Similarity 77.8%; Pred. No. 81;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLELVGVEE 13
Db 6 LLELVGVEE 14

RESULT 20
AT7A_MOUSE STANDARD; PRT; 1491 AA.
ID AT7A_MOUSE
```

```
AC 064430; Q64431; O35101; P97422;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COPPER-TRANSPORTING ATPASE 1 (EC 3.6.3.4) (COPPER PUMP 1) (MENKES
DE DISEASE-ASSOCIATED PROTEIN HOMOLOG).
GN ATP7A OR MNK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RC MEDLINE=94332144; PubMed=8054976;
RX Levinson B., Vulpe C., Elder B., Martin C., Verley F., Packman S.,
RA Gitschler J.; gene is the mouse homologue of the Menkes disease gene."
RA "The mottled gene is the mouse homologue of the Menkes disease gene."
RL Nat. Genet. 6:369-373(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR X SWISS WEBSTER, DL, AND BALB/C; TISSUE=Kidney, and Embryo;
RC MEDLINE=94332145; PubMed=8054977;
RX Mercer J.F.B., Grimes A., Ambrosini L., Lockhart P., Paynter J.A.,
RA Dierick H., Glover T.W.; "Mutations in the murine homologue of the Menkes gene in dappled and
RA "Mutations in the murine homologue of the Menkes gene in dappled and
RL blotchy mice."
RL Nat. Genet. 6:374-378(1994).
RN [3]
RP SEQUENCE FROM N.A. AND VARIANTS ARG-674 AND PRO-1381.
RC STRAIN=C3H; TISSUE=Placenta;
RC MEDLINE=98046561; PubMed=9385451;
RX Ohta Y., Shiraiishi N., Nishikimi M.;
RA "Occurrence of two missense mutations in Cu-ATPase of the macular
RA mouse, a Menkes disease model."
RL Biochem. Mol. Biol. Int. 43:913-918(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CBA X C3H;
RC MEDLINE=97358576; PubMed=9215672;
RX Grimes A., Hearn C.J., Lockhart P., Newgreen D.F., Mercer J.F.B.;
RA "Molecular basis of the brindled mouse mutant (Mo(br)): a murine model
RA of Menkes disease."
RL Hum. Mol. Genet. 6:1037-1042(1997).
CC -!- FUNCTION: MAY SUPPLY COPPER TO COPPER-REQUIRING PROTEINS WITHIN
CC THE SECRETORY PATHWAY. WHEN LOCALIZED IN THE TRANS-GOLGI NETWORK.
CC UNDER CONDITIONS OF ELEVATED EXTRACELLULAR COPPER, IT RELOCALIZED
CC TO THE PLASMA MEMBRANE WHERE IT FUNCTIONS IN THE EFFLUX OF COPPER
CC FROM CELLS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + CU(2+)(IN) = ADP + PHOSPHATE +
CC CU(2+)(OUT).
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONSTITUTIVELY
CC CYCLES BETWEEN THE TRANS-GOLGI NETWORK (TGN) AND THE PLASMA
CC MEMBRANE. PREDOMINANTLY FOUND IN THE TGN AND RELOCALIZED TO THE
CC PLASMA MEMBRANE IN RESPONSE TO ELEVATED COPPER LEVELS (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES EXCEPT LIVER. IN THE
CC KIDNEY, IT IS DETECTED IN THE PROXIMAL AND DISTAL TUBULES.
CC -!- DEVELOPMENTAL STAGE: WIDESPREAD EXPRESSED THROUGHOUT DEVELOPMENT.
CC DISEASE: DEFECTS IN ATP7A ARE ASSOCIATED WITH MOTTLED, AN X-LINKED
CC RECESSIVE CONDITION CHARACTERIZED BY MOTTLED PIGMENTATION OF THE
CC COAT, DEFECTS IN CONNECTIVE TISSUE AND NEONATAL OR FETAL DEATH. IT
CC IS DUE TO A DEFECT IN ABSORPTION AND TRANSPORT OF COPPER. THESE
CC MOTTLED MUTANTS EXHIBIT A DIVERSITY OF PHENOTYPES. TWO OF THESE
CC MUTANTS ARE CALLED BRINDLED AND BLOTCHY AND THEIR PHENOTYPES
CC RESEMBLE CLASSICAL MENKES DISEASE (MD) AND OCCIPITAL HORN SYNDROME
CC (OHS) IN HUMANS, RESPECTIVELY. OTHER MUTANTS ARE CALLED DAPPLED,
CC MOSAIC, TORTOISESHELL, PEWTER, ETC.
CC -!- MISCELLANEOUS: THE C-TERMINAL DI-LEUCINE, LEU-1478 ET LEU-1479, IS
CC AN ENDOCYTIC TARGETING SIGNAL WHICH FUNCTIONS IN RETRIEVING
CC RECYCLING FROM THE PLASMA MEMBRANE TO THE TGN. MUTATION OF THE DI-
CC LEUCINE SIGNAL RESULTS IN THE ACCUMULATION OF THE PROTEIN IN THE
```



CC PLASMA MEMBRANE (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY

CC (E1-E2 ATPASES). SUBFAMILY IB.

CC -!- SIMILARITY: CONTAINS 6 HEAVY-METAL-ASSOCIATED (HMA) DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U03434; AAA57445.1; -.

DR EMBL; U03736; AAB08487.1; -.

DR EMBL; AB007134; BAA22369.1; -.

DR EMBL; U71091; AAB37301.1; -.

DR HSSP; Q04656; 2AW0.

DR MGD; MGI:99400; Atp7a.

DR InterPro; IPR001757; -.

DR InterPro; IPR001877; -.

DR InterPro; IPR001934; -.

DR Pfam; PF00122; E1-E2 ATPase; 2.

DR Pfam; PF00403; HMA; 6.

DR PRINTS; PR00119; CATATPASE.

DR PRINTS; PR00942; CUATPASEI.

DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.

DR PROSITE; PS01047; HMA; 6.

KW Hydrolase; Copper transport; Transmembrane; Phosphorylation;

KW ATP-binding; Metal-binding; Magnesium; Copper; Repeat; Golgi stack;

KW Disease mutation; Multigene family.

FT DOMAIN 1 644

FT TRANSMEM 645 666

FT DOMAIN 667 705

FT TRANSMEM 706 725

FT DOMAIN 726 732

FT TRANSMEM 733 753

FT DOMAIN 754 772

FT TRANSMEM 773 793

FT DOMAIN 794 926

FT TRANSMEM 927 951

FT DOMAIN 951 980

FT TRANSMEM 981 1002

FT DOMAIN 1003 1347

FT TRANSMEM 1348 1365

FT DOMAIN 1366 1377

FT TRANSMEM 1377 1396

FT DOMAIN 1397 1491

FT DOMAIN 14 43

FT DOMAIN 177 206

FT DOMAIN 283 310

FT DOMAIN 383 412

FT DOMAIN 485 514

FT DOMAIN 561 590

FT DOMAIN 356 362

FT METAL 1292

FT METAL 1296

FT SITE 1478

FT MOD\_RES 1035

FT CARBOHYD 677

FT CARBOHYD 966

FT VARIANT 674 674

FT VARIANT 1381 1381

FT CONFLICT 44 44

FT CONFLICT 103 103

FT CONFLICT 172 172

FT CONFLICT 245 246

FT CONFLICT 445 445

FT CONFLICT 470 470

FT CONFLICT 515 515

FT CONFLICT 717 717

FT CONFLICT 770 770

FT CONFLICT 775 775

FT CONFLICT 775 775

FT CONFLICT 885 885 I -> T (IN REF. 2).

FT CONFLICT 1169 1169 Y -> H (IN REF. 2).

FT CONFLICT 1204 1204 A -> P (IN REF. 2 AND 4).

FT CONFLICT 1217 1217 I -> M (IN REF. 1).

FT CONFLICT 1253 1253 R -> Q (IN REF. 1).

SQ SEQUENCE 1491 AA; 161909 MW; B916EF9E2565247C CRC64;

Query Match 45.1%; Score 37; DB 1; Length 1491;

Best Local Similarity 53.8%; Pred. No. 1.8e+02;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLLLEVGVEEK 14

DB 143 VPDLSLDMGTOEK 155

II I I I I I I I I

RESULT 21

AT7A\_RAT

ID AT7A\_RAT STANDARD; PRT; 1492 AA.

AC P70705;

DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE COPPER-TRANSPORTING ATPASE 1 (EC 3.6.3.4) (COPPER PUMP 1) (MENKES

DE DISEASE-ASSOCIATED PROTEIN HOMOLOG).

GN ATP7A OR MNK.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Astocytes;

RX MEDLINE=98220710; PubMed=9562241;

RA Qian Y., Tiffany-Castiglioni E., Harris E.D.;

RT "Sequence of a Menkes-type Cu-transporting ATPase from rat C6 glioma

RT cells: comparison of the rat protein with other mammalian

RT Cu-transporting ATPases."

RL Mol. Cell. Biochem. 181:49-61(1998).

CC -!- FUNCTION: MAY SUPPLY COPPER TO COPPER-REQUIRING PROTEINS WITHIN

CC THE SECRETORY PATHWAY, WHEN LOCALIZED IN THE TRANS-GOLGI NETWORK.

CC UNDER CONDITIONS OF ELEVATED EXTRACELLULAR COPPER, IT RELOCALIZED

CC TO THE PLASMA MEMBRANE WHERE IT FUNCTIONS IN THE EFFLUX OF COPPER

CC FROM CELLS (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + CU(2+)(IN) = ADP + PHOSPHATE +

CC CU(2+)(OUT).

CC -!- SUBUNIT: MONOMER (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONSTITUTIVELY

CC CYCLES BETWEEN THE TRANS-GOLGI NETWORK (TGN) AND THE PLASMA

CC MEMBRANE. PREDOMINANTLY FOUND IN THE TGN AND RELOCALIZED TO THE

CC PLASMA MEMBRANE IN RESPONSE TO ELEVATED COPPER LEVELS (BY

CC SIMILARITY).

CC -!- MISCELLANEOUS: THE C-TERMINAL DI-LEUCINE, LEU-1479 ET LEU-1480, IS

CC AN ENDOCYTIC TARGETING SIGNAL WHICH FUNCTIONS IN RETRIEVING

CC RECYCLING FROM THE PLASMA MEMBRANE TO THE TGN. MUTATION OF THE DI-

CC LEUCINE SIGNAL RESULTS IN THE ACCUMULATION OF THE PROTEIN IN THE

CC PLASMA MEMBRANE (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY

CC (E1-E2 ATPASES), SUBFAMILY IB.

CC -!- SIMILARITY: CONTAINS 6 HEAVY-METAL-ASSOCIATED (HMA) DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U59245; AAB06393.1; -.

DR HSSP; Q04656; 2AW0.

DR InterPro; IPR001757; -.



```
DR InterPro: IPR001877;
DR InterPro: IPR001934;
DR Pfam: PF00122; E1-E2_ATPase; 2.
DR Pfam: PF00403; HMA; 6.
DR PRINTS; PRO0119; CATATPASEI.
DR PRINTS; PRO0942; CUATPASEI.
DR PROSITE; PS00154; ATPASE_E1_E2; 1.
DR PROSITE; PS01047; HMA; 6.
KW Hydrolase; Copper transport; Transmembrane; Phosphorylation;
KW ATP-binding; Metal-binding; Magnesium; Copper; Repeat; Golgi stack;
KW Multigene family.
FT DOMAIN 1 645
FT TRANSMEM 646 667
FT DOMAIN 668 706
FT TRANSMEM 707 726
FT DOMAIN 727 733
FT TRANSMEM 734 754
FT DOMAIN 755 773
FT TRANSMEM 774 794
FT DOMAIN 795 927
FT TRANSMEM 928 951
FT DOMAIN 952 981
FT TRANSMEM 982 1003
FT DOMAIN 1004 1348
FT TRANSMEM 1349 1366
FT DOMAIN 1367 1377
FT TRANSMEM 1378 1397
FT DOMAIN 1398 1492
FT TRANSMEM 14 43
FT DOMAIN 177 206
FT TRANSMEM 283 312
FT DOMAIN 383 412
FT TRANSMEM 486 515
FT DOMAIN 562 591
FT TRANSMEM 592 621
FT METAL 1293 1293
FT METAL 1297 1297
FT SITE 1479 1480
FT MOD_RES 1036 1036
FT CARBOHYD -678 678
FT SEQUENCE 1492 AA; 162091 MW; 34F75152B105AE9F CRC64;

Query Match 45.1%; Score 37; DB 1; Length 1492;
Best Local Similarity 53.8%; Pred. No. 1.8e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLLLEVGVVEK 14
DB 143 VPDLSLDMGTOEK 155

RESULT 22
DYHA_CHLRE STANDARD; PRT; 2405 AA.
AC Q39610;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DYNEIN ALPHA CHAIN, FLAGELLAR OUTER ARM (FRAGMENT).
GN ODA11 OR ODA-11.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21GR;
RX MEDLINE=94274778; PubMed=8006077;
RA Mitchell D.R.; Brown K.S.;
RT "Sequence analysis of the Chlamydomonas alpha and beta dynein heavy
chain genes."; 107:635-644(1994).
RL J. Cell Sci. 107:635-644(1994).
```

```
CC -!- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -!- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND
CC GAMMA). 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L26049; AAA57316.1;
CC Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
CC Coiled coil.
KW NON_TER 1
KW DOMAIN 120 193
KW DOMAIN 241 309
FT DOMAIN 695 723
FT DOMAIN 1514 1547
FT DOMAIN 1862 1882
FT DOMAIN 2029 2121
FT DOMAIN 2345 2374
FT NP_BIND 575 582
FT NP_BIND 1228 1235
FT NP_BIND 1576 1583
FT NON_TER 2405 2405
FT SEQUENCE 2405 AA; 273743 MW; 8471B2DFOFF8795A CRC64;

Query Match 45.1%; Score 37; DB 1; Length 2405;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 YVPPLLLLEV 9
DB 1071 YIPPTLLEM 1079

RESULT 23
Y045_MYCGE STANDARD; PRT; 483 AA.
AC P47291;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL LIPOPROTEIN MG045 PRECURSOR.
GN MG045.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M.; Gocayne J.D.; White O.; Adams M.D.; Clayton R.A.;
RA Fleischmann J.L.; Weidman J.F.; Small K.V.; Sandusky M.; Fuhrmann J.L.;
RA Fritchman D.T.; Utterback B.A.; Saudek D.M.; Phillips C.A.; Merrick J.M.;
RA Nguyen J.-F.; Dougherty B.A.; Bott K.F.; Hu P.-C.; Lucier J.C.;
RA Peterson S.N.; Smith H.O.; Hutchison C.A. III; Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 2-127 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N.; Hu P.-C.; Bott K.F.; Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
```

RT sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID  
CC ANCHOR (POTENTIAL).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U39684; AAC71261.1; -.  
DR EMBL; U02166; AAD12448.1; -.  
DR TIGR; MG045; -.  
DR InterPro; IPR000044; -.  
DR Pfam; PF02030; Lipoprotein\_8; 1.  
DR PRINTS; PRO0905; MYCMG045.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Hypothetical protein; Lipoprotein; Membrane; Signal.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 483 HYPOTHETICAL LIPOPROTEIN MG045.  
FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 483 AA; 54875 MW; 899D4CB60D78F69B CRC64;  
  
Query Match 44.5%; Score 36.5; DB 1; Length 483;  
Best Local Similarity 64.3%; Pred. No. 70;  
Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
  
QY 1 YVPPLLEVGVEEK 14  
|:|||||:|  
Db 36 YISPLLE-RVQEK 48  
  
RESULT 24  
Y850\_HAEIN STANDARD; PRT; 144 AA.  
AC P44060;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE HYPOTHETICAL PROTEIN HI0850.  
GN HI0850.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -!- SIMILARITY: TO E.COLI RSEC AND H.INFLUENZAE HI589.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U32766; AAC22507.1; -.  
DR TIGR; HI0850; -.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 76 96 POTENTIAL.  
FT TRANSMEM 105 125 POTENTIAL.  
SQ SEQUENCE 144 AA; 15841 MW; 564FC10BDE2CA4BD CRC64;  
  
Query Match 43.9%; Score 36; DB 1; Length 144;  
Best Local Similarity 54.5%; Pred. No. 24;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
  
QY 6 LLEVGVEEKFM 16  
::|:|:|:|  
Db 65 IVEIGLEKSM 75  
  
RESULT 25  
YIPA\_YEAST STANDARD; PRT; 248 AA.  
AC P53039;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE YIP1 PROTEIN.  
GN YIP1 OR YGR172C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HLR3;  
RA Yang X., Gallwitz D.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hebling U., Hofmann B., Delius H.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBBJ databases.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -!- SIMILARITY: TO S.POMBE SPCC61.04C.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X97342; CAA66031.1; -.  
DR EMBL; Z72957; CAA97198.1; -.  
DR SGD; S0003404; YIP1.  
KW Transmembrane.  
FT TRANSMEM 111 131 POTENTIAL.  
FT TRANSMEM 173 193 POTENTIAL.  
FT TRANSMEM 197 217 POTENTIAL.  
FT TRANSMEM 228 248 POTENTIAL.  
SQ SEQUENCE 248 AA; 27080 MW; 64372B4823A0BC5E CRC64;  
  
Query Match 43.9%; Score 36; DB 1; Length 248;  
Best Local Similarity 66.7%; Pred. No. 43;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 3 PPLLEVG 11  
|||:|:  
Db 71 PPLLEEIGI 79

```

RESULT 26
FHUF_ECOLI STANDARD; PRT; 262 AA.
AC P39405;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FERRIC IRON REDUCTASE PROTEIN FHUF.
GN FHUF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=99144129; PubMed=9990318;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99144129; PubMed=9990318;
RA Muller K., Matzanke B.F., Schunemann V., Trautwein A.X., Hantke K.;
RT "FhuF, an iron-regulated protein of Escherichia coli with a new type
RT of 2Fe-2S center.";
RL Eur. J. Biochem. 258:1001-1008(1998).
CC -!- FUNCTION: INVOLVED IN THE REDUCTION OF FERRIC IRON IN CYTOPLASMIC
CC -!- FERRIOXAMINE B.
CC -!- COFACTOR: BINDS A 2FE-2S CLUSTER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC; SOMETIME MEMBRANE-ASSOCIATED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U14003; AAA97266.1; -.
CC EMBL: AE000507; AAC77323.1; -.
CC Ecogene; EGI2595; fhuF.
CC Iron-sulfur.
CC METAL 244 244 IRON-SULFUR (2FE-2S).
CC METAL 245 245 IRON-SULFUR (2FE-2S).
CC METAL 256 256 IRON-SULFUR (2FE-2S).
CC METAL 259 259 IRON-SULFUR (2FE-2S).
CC SEQUENCE 262 AA; 30112 MW; 7708E776B1BE7783 CRC64;
DR
DR EMBL: U14003; AAA97266.1; -.
DR EMBL: AE000507; AAC77323.1; -.
DR Ecogene; EGI2595; fhuF.
KW Iron-sulfur.
FT METAL 244 244 IRON-SULFUR (2FE-2S).
FT METAL 245 245 IRON-SULFUR (2FE-2S).
FT METAL 256 256 IRON-SULFUR (2FE-2S).
FT METAL 259 259 IRON-SULFUR (2FE-2S).
FT SEQUENCE 262 AA; 30112 MW; 7708E776B1BE7783 CRC64;
SQ
Query Match 43.9%; Score 36; DB 1; Length 262;
Best Local Similarity 53.8%; Pred. NO. 45;
Matches 7; Conservative 3; Mismatches 3; Indels 3; Gaps 0;
QY 2 VPPLLLLEVGVEEK 14
Db 104 VPPLMLALLTQEK 116
RESULT 27
PIV6_ADE40 STANDARD; PRT; 267 AA.
AC P48309;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MINOR CAPSID PROTEIN PVI PRECURSOR.
GN PVI.
OS Human adenovirus type 40.

```

```

Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
NCBI_TaxID=28284;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DUGAN;
RX MEDLINE=94087748; PubMed=8263936;
RA Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT "The DNA sequence of adenovirus type 40.";
RL J. Mol. Biol. 234:1308-1316(1993).
CC -!- FUNCTION: MINOR CAPSID PROTEIN THAT MAY ACT AS A LINK BETWEEN THE
CC EXTERNAL CAPSID AND THE INTERNAL DNA-PROTEIN CORE.
CC -!- FUNCTION: THE C-TERMINAL DOMAIN MAY ACT AS A PROTEASE COFACTOR
CC WHICH ACTS AS AN ACTIVATOR OF THE ENZYME.
CC -!- SUBUNIT: THE PROTEASE COFACTOR IS A DISULFIDE-LINKED HOMODIMER
CC (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L19443; AAC13966.1; -.
CC Coat protein.
CC PROPEP 1 33 POTENTIAL.
CC CHAIN 34 256 MINOR CAPSID PROTEIN VI.
CC PEPTIDE 257 267 PROTEASE COFACTOR.
CC SITE 33 34 CLEAVAGE (BY ADENOVIRUS PROTEASE)
CC SITE 256 257 (BY SIMILARITY).
CC SITE 256 257 CLEAVAGE (BY ADENOVIRUS PROTEASE)
CC DISULFID 266 266 (BY SIMILARITY).
CC SEQUENCE 267 AA; 29129 MW; A634CADE3766F9C4 CRC64;
DR
DR EMBL: L19443; AAC13966.1; -.
KW Coat protein.
FT PROPEP 1 33 POTENTIAL.
FT CHAIN 34 256 MINOR CAPSID PROTEIN VI.
FT PEPTIDE 257 267 PROTEASE COFACTOR.
FT SITE 33 34 CLEAVAGE (BY ADENOVIRUS PROTEASE)
FT SITE 256 257 (BY SIMILARITY).
FT SITE 256 257 CLEAVAGE (BY ADENOVIRUS PROTEASE)
FT DISULFID 266 266 (BY SIMILARITY).
FT SEQUENCE 267 AA; 29129 MW; A634CADE3766F9C4 CRC64;
SQ
Query Match 43.9%; Score 36; DB 1; Length 267;
Best Local Similarity 66.7%; Pred. NO. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 3; Gaps 0;
QY 3 PPLLLLEVGVEEK 14
Db 122 PPKVEEVEVEEK 133
RESULT 28
KHSE_SYNY3 STANDARD; PRT; 306 AA.
AC P73646;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HOMOSERINE KINASE (EC 2.7.1.39) (HK).
GN THRB OR SLLI760.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8905231;
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-HOMOSERINE = ADP + O-PHOSPHO-L-
CC -!- HOMOSERINE.
CC -!- PATHWAY: THREONINE BIOSYNTHESIS.

```

```
CC -!- SIMILARITY: BELONGS TO THE GHMP KINASE FAMILY. HOMOSERINE
CC KINASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90908; BAA17691.1; -.
CC InterPro; IPR000870; -.
CC InterPro; IPR001745; -.
CC Pfam; PF00288; GHMP_kinases; 1.
CC PRINTS; PR00958; HOMSERKINASE.
CC PROSITE; PS00627; GHMP_KINASES_ATP; 1.
KW Threonine biosynthesis; Transferase; Kinase; ATP-binding.
FT NP_BIND 91 101 ATP (POTENTIAL).
SQ SEQUENCE 306 AA; 32427 MW; F1DFA3B3DEE45F36 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 306;
Best Local Similarity 54.5%; Pred. No. 53;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLLEVGVE 12
Db 79 IPPLKLEIDLE 89

RESULT 29
MAP3_SCHPO
ID MAP3_SCHPO STANDARD; PRT; 365 AA.
AC P31397;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHEROMONE M-FACTOR RECEPTOR.
GN MAP3 OR SPAC3F10.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST711;
RX MEDLINE=93109361; PubMed=8380233;
RA Tanaka K., Davey J., Imai Y., Yamamoto M.;
RT "Schizosaccharomyces pombe map3+ encodes the putative M-factor
RT receptor.";
RL Mol. Cell. Biol. 13:80-88(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE PEPTIDE PHEROMONE M-FACTOR, A MATING
CC FACTOR OF S.POMBE. PHEROMONE SIGNALING IS ESSENTIAL FOR INITIATION
CC OF MEIOSIS IN S.POMBE; M-FACTOR SIGNALING ALONE MAY BE SUFFICIENT.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- INDUCTION: BY AT LEAST 3 TYPES OF REGULATION: THE MATING-TYPE OF
CC THE CELL, NUTRITIONAL CONDITIONS AND PHEROMONE SIGNALING.
CC -!- SIMILARITY: BELONGS TO FAMILY 4 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
CC EMBL; D10933; BAA01727.1; -.
CC DR EMBL; Z69369; CAA93308.1; -.
CC DR PIR; A48105; A48105.
CC DR GCRDb; GCR_0565; -.
CC DR InterPro; IPR001499; -.
CC DR Pfam; PF02076; STE3; 2.
KW Transmembrane; G-protein coupled receptor; Pheromone response.
FT TRANSMEM 7 24 POTENTIAL.
FT TRANSMEM 31 54 POTENTIAL.
FT TRANSMEM 73 100 POTENTIAL.
FT TRANSMEM 116 133 POTENTIAL.
FT TRANSMEM 155 182 POTENTIAL.
FT TRANSMEM 204 226 POTENTIAL.
FT TRANSMEM 265 283 POTENTIAL.
SQ SEQUENCE 365 AA; 42498 MW; 306A022FA7DCFEF7 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 365;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPPLLLEVG 10
Db 161 YIPPCLLSFG 170

RESULT 30
DNAJ_BACSH
ID DNAJ_BACSH STANDARD; PRT; 368 AA.
AC O69269;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN DNAJ.
GN DNAJ.
OS Bacillus sphaericus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1421;
RN [1]
RP SEQUENCE FROM N.A..
RC STRAIN=1593;
RA Ahmad S., Selvapandiyan A., Gasbarri M., Bhatnagar R.K.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



DR	PROSITE; PS00637; DNAJ_CXXCXGXG; 1.
KW	Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.
FT	DOMAIN 5 65 J-DOMAIN.
FT	DOMAIN 72 97 GLY-RICH.
FT	REPEAT 138 145 CXXCXGXG MOTIF.
FT	REPEAT 155 162 CXXCXGXG MOTIF.
FT	REPEAT 181 188 CXXCXGXG MOTIF.
FT	REPEAT 195 202 CXXCXGXG MOTIF.
FT	METAL 138 138 ZINC 1 (BY SIMILARITY).
FT	METAL 141 141 ZINC 1 (BY SIMILARITY).
FT	METAL 155 155 ZINC 2 (BY SIMILARITY).
FT	METAL 158 158 ZINC 2 (BY SIMILARITY).
FT	METAL 181 181 ZINC 2 (BY SIMILARITY).
FT	METAL 184 184 ZINC 2 (BY SIMILARITY).
FT	METAL 195 195 ZINC 1 (BY SIMILARITY).
FT	METAL 198 198 ZINC 1 (BY SIMILARITY).
SQ	SEQUENCE 368 AA; 40692 MW; 6C69F02F52E9F1E6 CRC64;
Query Match 43.9%; Score 36; DB 1; Length 368;	
Best Local Similarity 53.3%; Pred. NO. 65;	
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;	
QY	1 YVPPLLEVGVEKF 15           :
Db	28 YHPDLNKEPGADEKF 42
RESULT	31
CATB_RHOOP	STANDARD; PRT; 373 AA.
ID	P95608;
AC	15-JUL-1999 (Rel. 38, Created)
DT	15-JUL-1999 (Rel. 38, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	MUCONATE CYCLOISOMERASE I (EC 5.5.1.1) (CIS,CIS-MUCONATE LACTONIZING ENZYME I) (MLE).
DE	ENZYME I) (MLE).
GN	CATB.
OS	Rhodococcus opacus.
OC	Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcu.
OX	NCBL_TaxID=37919;
RN	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP	[1]
RC	STRAIN=ICP;
RX	MEDLINE=97144521; Pubmed=8990288;
RA	Eulberg D.; Golovleva L.A.; Schloemann M.;
RT	"Characterization of catechol catabolic genes from Rhodococcus erythropolis TCP".
RL	J. Bacteriol. 179:370-381(1997).
CC	-1- CATALYTIC ACTIVITY: 2,5-DIHYDRO-5-OXOFURAN-2-ACETATE = CIS,CIS-HEXADIENEIOATE.
CC	-1- COFACTOR: REQUIRES A MANGANESE ION.
CC	-1- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUCONATE LACTONIZING ENZYME FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; X99622; CAA67934.1; -
DR	InterPro; IPR001354; -
DR	Pfam; PF01188; MR_MLE_1; 1.
DR	PROSITE; PS00908; MR_MLE_1; 1.
DR	PROSITE; PS00909; MR_MLE_2; 1.
KW	Aromatic hydrocarbons catabolism; Isomerase; Manganese.
FT	ACT_SITE 169 169 BY SIMILARITY.
FT	METAL 198 198 MANGANESE (BY SIMILARITY).
FT	METAL 224 224 MANGANESE (BY SIMILARITY).

Query Match 43.9%; Score 36; DB 1; Length 373; Best Local Similarity 46.2%; Pred. No. 66; Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;																												
QY	1	YVPPLLEVGVEE	13																									
Db	277	HLPPVLLAAGVPQ	289																									
RESULT 33																												
ID	DNAJ_LISMO	STANDARD;	PRT;	377	AA.																							
AC	Q9S5A3;																											
DT	01-OCT-2000	(Rel. 40, Created)																										
DT	01-OCT-2000	(Rel. 40, Last sequence update)																										
DT	01-OCT-2000	(Rel. 40, Last annotation update)																										
DE	CHAPERONE PROTEIN	DNAJ.																										
GN	DNAJ.																											
OS	Listeria monocytogenes.																											
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;																											
OC	Bacillus/Staphylococcus group; Listeria.																											
OX	NCBI_TaxID=1639;																											
RN	[1]																											
RP	SEQUENCE FROM N.A.																											
RX	STRAIN=10403S;																											
RX	MEDLINE=20163771; PubMed=10701836;																											
RA	Hanawa T., Kai M., Kamiya S., Yamamoto T.;																											
RT	"Cloning, sequencing, and transcriptional analysis of the dnaK heat																											
RT	shock operon of Listeria monocytogenes.";																											
RL	Cell Stress Chaperones 5:21-29(2000).																											
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,																											
CC	THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).																											
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).																											
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).																											
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.																											
CC	-!- SIMILARITY: CONTAINS 1 J DOMAIN.																											
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.																											
CC	-----																											
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration																											
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -																											
CC	the European Bioinformatics Institute. There are no restrictions on its																											
CC	use by non-profit institutions as long as its content is in no way																											
CC	modified and this statement is not removed. Usage by and for commercial																											
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>																											
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).																											
CC	-----																											
DR	EMBL; AB023064; BAA82790.1; -.																											
DR	InterPro; IPR001305; -.																											
DR	InterPro; IPR001623; -.																											
DR	InterPro; IPR002939; -.																											
DR	InterPro; IPR003095; -.																											
DR	Pfam; PF00226; DnaJ; 1.																											
DR	Pfam; PF00684; DnaJ_CXXCXGXG; 1.																											
DR	Pfam; PF01556; DnaJ_C; 1.																											
DR	PRINTS; PR00625; DNAJPROTEIN.																											
DR	PROSITE; PS00636; DNAJ_1; 1.																											
DR	PROSITE; PS50076; DNAJ_2; 1.																											
DR	PROSITE; PS00637; DNAJ_CXXCXGXG; 1.																											
KW	Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.																											
FT	DOMAIN 5 69																											
FT	DOMAIN 76 117																											
FT	REPEAT 147 154																											
FT	REPEAT 164 171																											
FT	REPEAT 190 197																											
FT	REPEAT 204 211																											
FT	METAL 147 147																											
FT	METAL 150 150																											
FT	METAL 164 164																											
FT	METAL 167 167																											
FT	METAL 190 190																											
FT	METAL 193 193																											
FT	METAL 204 204																											
FT	METAL 207 207																											
-----																												

Query Match 43.9%; Score 36; DB 1; Length 377; Best Local Similarity 46.7%; Pred. No. 66; Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;																												
QY	1	YVPPLLEVGVEEK	15																									
Db	32	YHPDINKEAGADEKF	46																									
RESULT 34																												
ID	DNAJ_LACLA	STANDARD;	PRT;	379	AA.																							
AC	P35514;																											
DT	01-JUN-1994	(Rel. 29, Created)																										
DT	01-JUN-1994	(Rel. 29, Last sequence update)																										
DT	01-OCT-2000	(Rel. 40, Last annotation update)																										
DE	CHAPERONE PROTEIN	DNAJ.																										
GN	DNAJ.																											
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).																											
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;																											
OC	Lactococcus.																											
OX	NCBI_TaxID=1360;																											
RN	[1]																											
RP	SEQUENCE FROM N.A.																											
RX	STRAIN=NIZO R5;																											
RX	MEDLINE=93194788; PubMed=8449872;																											
RA	van Asseldonk M., Simons A., Visser H., de Vos W.M., Simons G.;																											
RT	"Cloning, nucleotide sequence, and regulatory analysis of the																											
RT	Lactococcus lactis dnaJ gene.";																											
RL	J. Bacteriol. 175:1637-1644(1993).																											
CC	-!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,																											
CC	THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).																											
CC	-!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).																											
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).																											
CC	-!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.																											
CC	-!- SIMILARITY: CONTAINS 1 J DOMAIN.																											
CC	-!- SIMILARITY: CONTAINS 1 CR DOMAIN.																											
CC	-----																											
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration																											
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -																											
CC	the European Bioinformatics Institute. There are no restrictions on its																											
CC	use by non-profit institutions as long as its content is in no way																											
CC	modified and this statement is not removed. Usage by and for commercial																											
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>																											
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).																											
CC	-----																											
DR	EMBL; M99413; -; NOT_ANNOTATED_CDS.																											
DR	PIR; A47079; A47079.																											
DR	HSSP; P08622; 1XBL.																											
DR	InterPro; IPR001305; -.																											
DR	InterPro; IPR001623; -.																											
DR	InterPro; IPR002939; -.																											
DR	InterPro; IPR003095; -.																											
DR	Pfam; PF00226; DnaJ; 1.																											
DR	Pfam; PF01556; DnaJ_C; 1.																											
DR	Pfam; PF00684; DnaJ_CXXCXGXG; 1.																											
DR	PRINTS; PR00625; DNAJPROTEIN.																											
DR	PROSITE; PS00636; DNAJ_1; 1.																											
DR	PROSITE; PS50076; DNAJ_2; 1.																											
DR	PROSITE; PS00637; DNAJ_CXXCXGXG; 1.																											
KW	Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.																											
FT	DOMAIN 5 69																											
FT	DOMAIN 72 113																											
FT	REPEAT 154 161																											
FT	REPEAT 171 178																											
FT	REPEAT 197 204																											
FT	REPEAT 211 218																											

FT METAL 174 174 ZINC 2 (BY SIMILARITY).  
FT METAL 197 197 ZINC 2 (BY SIMILARITY).  
FT METAL 200 200 ZINC 2 (BY SIMILARITY).  
FT METAL 211 211 ZINC 1 (BY SIMILARITY).  
FT METAL 214 214 ZINC 1 (BY SIMILARITY).  
SQ SEQUENCE 379 AA; 40786 MW; 941C7E883433969B CRC64;  
  
Query Match 43.9%; Score 36; DB 1; Length 379;  
Best Local Similarity 53.3%; Pred. NO. 67;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 YVPPLLELVGVVEEK 15  
DB 32 YHPDLNKEEGAEKY 46  
  
RESULT 35  
YSH8\_CAEEL STANDARD; PRT; 393 AA.  
AC Q09949;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 44.8 KDA PROTEIN F12A10.8 IN CHROMOSOME II.  
GN F12A10.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Geisel C.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U28731; AAA68301.1;  
DR WormPep; F12A10.8; CE01911.  
DR Hypothetical protein.  
KW DOMAIN 157 161 POLY-GLU.  
FT SEQUENCE 393 AA; 44843 MW; 40ECLA3B84922C6 CRC64;  
SQ  
  
Query Match 43.9%; Score 36; DB 1; Length 393;  
Best Local Similarity 63.6%; Pred. NO. 69;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 PLLLEVGVVEEK 14  
DB 323 PYYLEIGVEEK 333  
  
RESULT 36  
SACB\_ERWAM STANDARD; PRT; 415 AA.  
AC Q46654;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE LEVANSUCRASE (EC 2.4.1.10) (BETA-D-FRUCTOFURANOSYL TRANSFERASE)  
DE (SUCROSE 6-FRUCTOSYL TRANSFERASE).  
GN LSC.  
OS Erwinia amylovora.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Erwinia.

OX NCBI\_TaxID=552;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EA7/74;  
RA Geier G., Geider K.K.; influence on virulence of the levansucrase gene  
RT from the fireblight pathogen Erwinia amylovora.";  
RL Physiol. Mol. Plant Pathol. 42:387-404(1993).  
CC CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -  
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN  
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X75079; CAA52972.1;  
DR Transferase; Glycosyltransferase.  
KW SEQUENCE 415 AA; 46408 MW; 4FF564F6E0607FEB CRC64;  
SQ  
  
Query Match 43.9%; Score 36; DB 1; Length 415;  
Best Local Similarity 46.2%; Pred. NO. 73;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 VPPLLLELVGVVEEK 14  
DB 273 LPPLITAVGVNDQ 285  
  
RESULT 37  
SPAK\_HUMAN STANDARD; PRT; 547 AA.  
ID Q9UEW8; Q9UE84; OL4774;  
AC Q9UEW8; Q9UE84; OL4774;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE STE20/SPSL-RELATED PROLINE-ALANINE RICH PROTEIN KINASE (EC 2.7.1.-)  
DE (STE-20 RELATED KINASE) (DCHT).  
GN SPAK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20438689; PubMed=10980603;  
RA Johnston A.M., Naselli G., Genez L.J., Martin R.M., Harrison L.C.,  
RA de Aizpurua H.J.;  
RT "Spak, a STE20/SPSL-related kinase that activates the p38 pathway.";  
RL Oncogene 19:4290-4297(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Melnick M.B., Pettit M., Perrimon N., Comb M.J.;  
RT "New human member of the Ste20 family.";  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 281-547 FROM N.A.  
RC TISSUE=Testis;  
RA Baytel D., Don J.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: MAY ACT AS A MEDIATOR OF STRESS-ACTIVATED SIGNALS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (WHEN CASPASE-  
CC CLEAVED) (PROBABLE).  
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN AND PANCREAS  
CC FOLLOWED BY HEART, LUNG, KIDNEY, SKELETAL MUSCLE, LIVER, PLACENTA

CC AND TESTIS.  
CC -!- DOMAIN: PAPA BOX (PROLINE-ALANINE REPEATS) MAY TARGET THE KINASE  
CC TO A SPECIFIC SUBCELLULAR LOCATION BY FACILITATING INTERACTION  
CC WITH INTRACELLULAR PROTEINS SUCH AS ACTIN OR ACTIN-LIKE PROTEINS.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC STE20 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF099989; AAC72238.1; --  
CC EMBL; AF030403; AAD01901.1; --  
CC EMBL; AF017635; AAB70552.1; --  
CC HSP; P00518; 1PHK.  
CC InterPro; IPR000719; --  
CC Pfam; PF00069; pkinase; 1.  
CC PROSITE; PS00108; PROTEIN\_KINASE\_ST; FALSE\_NEG.  
CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
CC PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW transferase; Serine/threonine-protein kinase; ATP-binding.  
FT DOMAIN 14 55 PRO/ALA-RICH.  
FT NP\_BIND 65 339 PROTEIN KINASE.  
FT BINDING 71 79 ATP (BY SIMILARITY).  
FT ACT\_SITE 94 94 ATP (BY SIMILARITY).  
FT DOMAIN 194 194 BY SIMILARITY.  
FT SITE 362 368 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT CONFLICT 389 393 CASPASE CLEAVAGE RELATED SITE.  
FT CONFLICT 175 175 L -> F (IN REF. 1).  
SQ SEQUENCE 547 AA; 59642 MW; DE1C71CD46A79986 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 547;  
Best Local Similarity 61.5%; Pred. No. 98;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEKFM 16  
| | | | |  
Db 292 PPTLETGVEDKEM 304

RESULT 38  
SPAK\_RAT  
ID SPK\_RAT STANDARD; PRT; 553 AA.  
AC O88506; 070541;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE STE20/SPS1-RELATED PROLINE-ALANINE RICH PROTEIN KINASE (EC 2.7.1.1-)  
DE (STE-20 RELATED KINASE) (PANCREATIC SERINE/THREONINE KINASE) (PS/TK)  
DE (PSTK1).  
GN SPK OR PASK.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=98342058; PubMed=9675032;  
RA Ushiro H., Tsutsumi T., Suzuki K., Kayahara T., Nakano K.;  
RT "Molecular cloning and characterization of a novel Ste20-related  
RT protein kinase enriched in neurons and transporting epithelia.";  
RL Arch. Biochem. Biophys. 355:233-240(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Pancreas;  
RX MEDLINE=20445994; PubMed=10990492;  
RA Miao N., Fung B., Sanchez R., Lydon J., Barker D., Pang K.;

RT "Isolation and expression of PASK, a serine/threonine kinase, during  
RT rat embryonic development, with special emphasis on the pancreas.";  
RL J. Histochem. Cytochem. 48:1391-1400(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Insulinoma;  
RX MEDLINE=20438689; PubMed=10980603;  
RA Johnston A.M., Naselli G., Genez L.J., Martin R.M., Harrison L.C.,  
RA de Aizpurua H.J.;  
RT "SPAK, a STE20/SPS1-related kinase that activates the p38 pathway.";  
RL Oncogene 19:4290-4297(2000).  
CC -!- FUNCTION: MAY ACT AS A MEDIATOR OF STRESS-ACTIVATED SIGNALS.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (WHEN CASPASE-  
CC CLEAVED) (PROBABLE).  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTES FOLLOWED BY  
CC PANCREAS, KIDNEY, HEART AND BRAIN. NOT EXPRESSED IN SKELETAL  
CC MUSCLE, LIVER, LUNG AND SPLEEN.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE EARLY GUT AND PANCREATIC  
CC EPITHELIUM, AT E15 DAY LOCALIZED TO CELLS THAT WILL EVENTUALLY  
CC BECOME EXOCRINE. EXPRESSED IN CHOROID PLEXUS, DEVELOPING  
CC MYOCARDIUM, PANCREATIC EPITHELIUM AND DORSAL ROOT GANGLIA.  
CC -!- DOMAIN: PAPA BOX (PROLINE-ALANINE REPEATS) MAY TARGET THE KINASE  
CC TO A SPECIFIC SUBCELLULAR LOCATION BY FACILITATING INTERACTION  
CC WITH INTRACELLULAR PROTEINS SUCH AS ACTIN OR ACTIN-LIKE PROTEINS.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC STE20 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D88190; BAA26000.1; --  
CC EMBL; AF068261; AAC23501.1; --  
CC EMBL; AF099990; AAC72239.1; --  
CC HSP; P00518; 1PHK.  
CC InterPro; IPR000719; --  
CC Pfam; PF00069; pkinase; 1.  
CC PROSITE; PS00108; PROTEIN\_KINASE\_ST; FALSE\_NEG.  
CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
CC PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW transferase; Serine/threonine-protein kinase; ATP-binding.  
FT DOMAIN 14 62 PRO/ALA-RICH.  
FT DOMAIN 72 346 PROTEIN KINASE.  
FT NP\_BIND 78 86 ATP (BY SIMILARITY).  
FT BINDING 101 101 ATP (BY SIMILARITY).  
FT ACT\_SITE 201 201 BY SIMILARITY.  
FT DOMAIN 22 30 POLY-ALA.  
FT DOMAIN 369 375 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT SITE 396 400 CASPASE CLEAVAGE RELATED SITE.  
FT CONFLICT 11 11 V -> I (IN REF. 2).  
FT CONFLICT 253 253 M -> I (IN REF. 2).  
FT CONFLICT 403 403 E -> Q (IN REF. 2).  
SQ SEQUENCE 553 AA; 60050 MW; 8BC1458AEA6F90C8 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 553;  
Best Local Similarity 61.5%; Pred. No. 99;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEKFM 16  
| | | | |  
Db 299 PPTLETGVEDKEM 311

RESULT 39  
SPAK\_MOUSE  
ID SPK\_MOUSE STANDARD; PRT; 556 AA.  
AC Q921W9;  
DT 01-OCT-2000 (Rel. 40, Created)



DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE STE20/SPS1-RELATED PROLINE-ALANINE RICH PROTEIN KINASE (EC 2.7.1.1-)  
DE (STE-20 RELATED KINASE)  
GN SPAN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=2043669; PubMed=10980603;  
RA Johnston A.M., Naselli G., Genez L.J., Martin R.M., Harrison L.C.,  
RA de Aizpurua H.J.;  
RL "SPAK, a STE20/SPS1-related kinase that activates the p38 pathway."  
RL Oncogene 19:4290-4297(2000).  
CC -!- FUNCTION: MAY ACT AS A MEDIATOR OF STRESS-ACTIVATED SIGNALS.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (WHEN CASPASE-  
CC CLEAVED) (PROBABLE).  
CC -!- DOMAIN: PAPA BOX (PROLINE-ALANINE REPEATS) MAY TARGET THE KINASE  
CC TO A SPECIFIC SUBCELLULAR LOCATION BY FACILITATING INTERACTION  
CC WITH INTRACELLULAR PROTEINS SUCH AS ACTIN OR ACTIN-LIKE PROTEINS.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC STE20 SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF099988; AAC72237.1;  
CC HSP; P00518; IPIK.  
CC MGD; MGI:1858416; Spak.  
CC InterPro; IPR000719;  
CC Pfam; PF00069; pkinase; 1.  
CC PROSITE; PS00108; PROTEIN\_KINASE\_ATP; 1.  
CC PROSITE; PS00107; PROTEIN\_KINASE\_DOM; 1.  
CC PROSITE; PS50011; PROTEIN-threonine-protein kinase; ATP-binding.  
CC Transferrase; Serine/threonine-protein kinase; ATP-binding.  
CC DOMAIN 17 65 PRO/ALA-RICH.  
CC DOMAIN 75 349 PROTEIN KINASE.  
CC NP\_BIND 81 89 ATP (BY SIMILARITY).  
CC BINDING 104 104 ATP (BY SIMILARITY).  
CC ACT\_SITE 204 204 BY SIMILARITY.  
CC DOMAIN 22 31 POLY-ALA.  
CC DOMAIN 372 378 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC SITE 399 403 CASPASE CLEAVAGE RELATED SITE.  
CC SEQUENCE 556 AA; 60319 MW; 66085A90554311D8 CRC64;  
Query Match 43.98; Score 36; DB 1; Length 556;  
Best Local Similarity 61.58; Pred. NO. 99;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 4 PLLLEGVGVEEKFM 16  
DB 302 PPTLETGVEDKEM 314  
RESULT 40  
ID IUCC\_ECOLI STANDARD; PRT; 580 AA.  
AC Q47318;  
DT 30-MAY-2000 (Rel. 39, Created)  
DE 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE AEROBACTIN SIDEROPHORE BIOSYNTHESIS PROTEIN IUCC.  
GN IUCC.  
OS Escherichia coli.

OG Plasmid IncFI ColV3-K30.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE=94210503; PubMed=8003107;  
RA Martinez J.L., Herrero M., de Lorenzo V.;  
RA "The organization of intercalistronic regions of the aerobactin operon  
RT of pColV-K30 may account for the differential expression of the  
RT iucABCD iuta genes."  
RL J. Mol. Biol. 238:288-293(1994).  
CC -!- PATHWAY: AEROBACTIN SIDEROPHORE BIOSYNTHESIS.  
CC -!- SIMILARITY: BELONGS TO THE IUCC FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X76100; CAA53709.1;  
CC Iron transport; Plasmid.  
CC SEQUENCE 580 AA; 66845 MW; 406FB857BB367F65 CRC64;  
Query Match 43.98; Score 36; DB 1; Length 580;  
Best Local Similarity 35.78; Pred. NO. 1e+02;  
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 1 YVPPLLLEVGVEEK 14  
DB 491 FISPLMVRGLGPER 504  
RESULT 41  
ID GLMS\_CHLPN STANDARD; PRT; 608 AA.  
AC Q9Z6U0; Q9JRV0;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE (ISOMERIZING)  
DE (EC 2.6.1.16) (HEXOSEPHOSPHATE AMINOTRANSFERASE) (D-FRUCTOSE-6-  
DE PHOSPHATE AMIDOTRANSFERASE) (GFAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE  
DE AMIDOTRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).  
GN GLMS OR CPN0968 OR CP0892.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Unayam L.A., Utterback T.,  
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,  
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39."  
RL Nucleic Acids Res. 28:1397-1406(2000).



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X00415; CAA25120.1; -  
EMBL; X64592; -; NOT ANNOTATED\_CDS.  
EMBL; U18997; AAA58137.1; -  
EMBL; AE000410; AAC76365.1; -  
EMBL; V00356; CAA23651.1; -  
EMBL; X65735; CAA46645.1; -  
PIR; JCI1426; JCI1426.  
PIR; A28513; EFECG.  
PIR; S24928; S24928.  
HSSP; P13551; IELO.  
SWISS-2DPAGE; P02996; COLI.  
ECODBASE; D084.0; 6TH EDITION.  
ECOGene; EG10360; fusa.  
InterPro; IPR000640; -  
InterPro; IPR000795; -  
Pfam; PF00679; EFG\_C; 1.  
Pfam; PF00009; GTP\_EFTU; 1.  
PRINTS; PR00315; ELONGATNECT.  
PROSITE; PS00301; EFATOR\_GTP; 1.  
Elongation factor; Protein biosynthesis; GTP-binding.  
INIT\_MET 0 0  
NP\_BIND 16 23 GTP (BY SIMILARITY).  
NP\_BIND 87 91 GTP (BY SIMILARITY).  
NP\_BIND 141 144 GTP (BY SIMILARITY).  
NP\_BIND 295 296 NG->DC (IN REF. 5).  
CONFLICT 299 301 MISSING (IN REF. 5).  
CONFLICT 395 395 T->C (IN REF. 5).  
CONFLICT 575 575 I->V (IN REF. 5 AND 6).  
CONFLICT 583 583 H->K (IN REF. 5 AND 6).  
CONFLICT 593 593 K->Q (IN REF. 5 AND 6).  
CONFLICT 625 625 E->Q (IN REF. 5 AND 6).  
CONFLICT 645 645 E->Q (IN REF. 5 AND 6).  
CONFLICT 656 656 E->EQ (IN REF. 5 AND 6).  
CONFLICT 661 661 E->EQ (IN REF. 5 AND 6).  
SEQUENCE 703 AA; 77450 MW; 4335C54C941F9103 CRC64;  
Query Match 43.9%; Score 36; DB 1; Length 703;  
Best Local Similarity 58.3%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 4 PLLLEVGVEEF 15  
DB 167 PLQLAIGAEHF 178  
RESULT 43  
ACPH\_HUMAN STANDARD; PRT; 732 AA.  
ID ACPH\_HUMAN  
AC P13798;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ACYLAMINO-ACID-RELEASES ENZYME (EC 3.4.19.1) (ACYL-PEPTIDE HYDROLASE)  
DE (APH) (ACYLAMINOACYL-PEPTIDASE) (DNF15S2 PROTEIN).  
GN APEH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Mitta M., Ohnogi H., Mizutani S., Kato I., Sakiyama F., Tsunasawa S.;  
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
[2]  
RN PRELIMINARY SEQUENCE OF 102-732 FROM N.A.  
RX MEDLINE=89233127; PubMed=2565880;  
RA Naylor S.L., Marshall A., Hensel C., Martinez P.F., Holley B.,

Sakaguchi A.Y.;  
"The DNF15S2 locus at 3p21 is transcribed in normal lung and small cell lung cancer.";  
Genomics 4:355-361(1989).  
[3]  
FUNCTION.  
MEDLINE=91319413; PubMed=1861871;  
ERLANDSSON R., Boldog F., Persson B., Zabarovsky E.R., Allikmets R.L., Sumegi J., Klein G., Joernvall H.;  
"The gene from the short arm of chromosome 3, at D3F15S2, frequently deleted in renal cell carcinoma, encodes acylpeptide hydrolase.";  
Oncogene 6:1293-1295(1991).  
[4]  
FUNCTION.  
MEDLINE=91172778; PubMed=2006156;  
JONES W.M., Scaloni A., Bossa F., Popowicz A.M., Schneewind O., Manning J.M.;  
"Genetic relationship between acylpeptide hydrolase and acylase, two hydrolytic enzymes with similar binding but different catalytic specificities.";  
Proc. Natl. Acad. Sci. U.S.A. 88:2194-2198(1991).  
[5]  
X-RAY CRYSTALLOGRAPHY.  
MEDLINE=94016571; PubMed=8411161;  
FEESSE M., Scaloni A., Jones W.M., Mannig J.M., Remington S.J.;  
"Crystallization and preliminary x-ray studies of human erythrocyte acylpeptide hydrolase.";  
J. Mol. Biol. 233:546-549(1993).  
-1- FUNCTION: THIS ENZYME CATALYZES THE HYDROLYSIS OF THE AMINO-TERMINAL PEPTIDE BOND OF AN N-ACETYLATED PEPTIDE TO GENERATE AN N-ACETYLATED AA AND A PEPTIDE WITH A FREE AMINO-TERMINUS.  
-1- IT PREFERENTIALLY CLEAVES OFF AC-ALA, AC-MET AND AC-SER.  
-1- CATALYTIC ACTIVITY: ACYLAMINOACYL-PEPTIDE + H(2)O = ACYLAMINO ACID + PEPTIDE.  
-1- SUBUNIT: HOMOTETRAMER.  
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9C; ALSO KNOWN AS THE PROLYL OLIGOPEPTIDASE FAMILY.  
-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY DUE TO FRAMESHIFT ERRORS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D38441; BAA07476.1; -  
EMBL; J03068; AAA35769.1; ALT\_FRAME.  
MEROPS; S09.004; -  
MIM; 102645; -  
InterPro; IPR001375; -  
InterPro; IPR002471; -  
Pfam; PF00326; Peptidase\_S9; 1.  
PROSITE; PS00708; PRO\_ENDOPEP\_SER; 1.  
Hydrolase; Acetylation.  
MOD\_RES 1 587  
ACT\_SITE 587 587  
ACT\_SITE 675 675  
ACT\_SITE 707 707  
SEQUENCE 732 AA; 81292 MW; 3842FB353CE78A40 CRC64;  
ACETYLATION (BY SIMILARITY).  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
CHARGE RELAY SYSTEM (BY SIMILARITY).

Query Match 43.9%; Score 36; DB 1; Length 732;  
Best Local Similarity 44.4%; Pred. No. 1.3e+02;  
Matches 8; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

QY 1 YVP-----PLLLEVGVEEF 14  
DB 659 YIPQVKTPLLMLGQEDR 676

```
RESULT 44
DPO1_HAEIN
ID DPO1_HAEIN STANDARD; PRT; 930 AA.
AC P43741;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE DNA POLYMERASE I (EC 2.7.7.7) (POL I).
GN POLA OR HI0856.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32767; AAC22515.1; -.
DR HSSP; P00582; 1KRP.
DR TIGR; HI0856; -.
DR InterPro; IPR001098; -.
DR InterPro; IPR002298; -.
DR InterPro; IPR002421; -.
DR InterPro; IPR002562; -.
DR Pfam; PF01612; 3_5_exonuclease; 1.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00476; DNA_pol_A; 1.
DR PRINTS; PR00868; DNAPOLI.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydrolase; Exonuclease; DNA-binding.
SQ SEQUENCE 930 AA; 103740 MW; 226654BB7CFF730B CRC64;
```

```
Query Match 43.9%; Score 36; DB 1; Length 930;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLLLEVGVVEKF 15
| | | | | | | |
Db 913 VVPLIVEVGVGQNW 926
```

```
RESULT 45
INVA_YERPS
ID INVA_YERPS STANDARD; PRT; 985 AA.
AC P11922;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE INVASIN.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87301720; PubMed=3304658;
RA Isberg R.R., Voorhis D.L., Falkow S.;
RT "Identification of invasin: a protein that allows enteric bacteria to
RT penetrate cultured mammalian cells.";
RL Cell 50:769-778(1987).
RN [2]
RP INTEGRIN-BINDING DOMAIN.
RX MEDLINE=90269235; PubMed=1693333;
RA Leong J.M., Fournier R.S., Isberg R.R.;
RT "Identification of the integrin binding domain of the Yersinia
RT pseudotuberculosis invasin protein.";
RL EMBO J. 9:1979-1989(1990).
CC -!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS.
CC -!- SUBCELLULAR LOCATION: OUTER SURFACE.
CC -!- SIMILARITY: BELONGS TO THE EAE/INVASIN FAMILY.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-17 OR MET-19 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17448; AAA27633.1; -.
DR EMBL; M17448; AAA27632.1; ALT_INIT.
DR EMBL; M17448; AAA27634.1; ALT_INIT.
DR EMBL; M17448; AAA27635.1; ALT_INIT.
DR PIR; A29646; A29646.
DR PIR; S12543; S12543.
DR DOMAIN 795 985 INTEGRIN-BINDING.
SQ SEQUENCE 985 AA; 106627 MW; C875941B24BD35EE CRC64;
```

```
Query Match 43.9%; Score 36; DB 1; Length 985;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLLEVGVVEEK 14
| | | | | | | |
Db 320 YTPVPLLTGVDQR 333
```

Search completed: June 28, 2001, 11:57:22  
Job time: 263 sec



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 72.61 Seconds  
(without alignments)  
29.154 Million cell updates/sec

Title: US-09-439-313-554  
Perfect score: 82  
Sequence: 1 YVPPLLLEVGEVKFM 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SPTREMBL\_16:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	59.8	599	5 Q9VSV1	Q9vsv1 drosophila
2	46	56.1	280	2 Q9RKH8	Q9rkh8 streptomyce
3	45	54.9	299	2 Q56171	Q56171 streptomyce
4	43	52.4	351	10 Q9LSL1	Q9lsl1 arabidopsis
5	42	51.2	230	1 Q9HLZ9	Q9hlz9 thermoplasm
6	42	51.2	316	2 Q9FC55	Q9fc55 streptomyce
7	42	51.2	367	2 Q9WYC6	Q9wyc6 thermotoga
8	42	51.2	449	2 Q9X3R1	Q9x3r1 pseudomonas
9	42	51.2	528	5 O16518	O16518 caenorhabdi
10	42	51.2	700	2 Q9I338	Q9i338 pseudomonas
11	41	50.0	260	2 Q54386	Q54386 streptomyce
12	41	50.0	309	2 Q9I217	Q9i217 pseudomonas
13	41	50.0	405	2 Q9X6R2	Q9x6r2 pseudomonas
14	41	50.0	436	2 O67814	O67814 aquifex aeo
15	41	50.0	3472	1 O74056	O74056 cenarchaeum
16	40	48.8	108	3 Q08504	Q08504 saccharomyc
17	40	48.8	136	2 Q56530	Q56530 unidentified
18	40	48.8	157	4 Q9NW52	Q9nw52 homo sapien
19	40	48.8	243	1 O28678	O28678 archaeoglob

20	40	48.8	261	5 Q9GVK5	Q9gvk5 leishmania
21	40	48.8	348	1 Q9Y8P8	Q9y8p8 aeropyrum p
22	40	48.8	407	5 Q9N389	Q9n389 caenorhabdi
23	40	48.8	437	10 Q96544	Q96544 beta vulgar
24	40	48.8	687	2 Q9JXQ5	Q9jxq5 neisseria m
25	40	48.8	687	2 Q9JW67	Q9jw67 neisseria m
26	40	48.8	737	2 O51274	O51274 borrelia bu
27	40	48.8	916	2 Q9LCT3	Q9lct3 pseudomonas
28	39	47.6	162	2 Q9K5S0	Q9k5s0 bacillus ha
29	39	47.6	286	5 Q9UOL8	Q9uol8 plasmodium
30	39	47.6	299	4 O95685	O95685 homo sapien
31	39	47.6	307	2 P72208	P72208 prochloroco
32	39	47.6	326	2 Q9KXR2	Q9kxr2 streptomyce
33	39	47.6	333	3 Q9P727	Q9p727 neurospora
34	39	47.6	419	5 P91430	P91430 caenorhabdi
35	39	47.6	531	5 Q21603	Q21603 caenorhabdi
36	39	47.6	570	10 Q9XEE4	Q9xee4 arabidopsis
37	39	47.6	606	2 Q9PLA4	Q9pla4 chlamydia m
38	39	47.6	674	4 Q9H6U0	Q9h6u0 homo sapien
39	39	47.6	800	5 Q9XWL9	Q9xwl9 caenorhabdi
40	39	47.6	1148	4 Q9H6W7	Q9h6w7 homo sapien
41	39	47.6	1963	10 Q9LXT9	Q9lxt9 arabidopsis
42	39	47.6	1972	10 Q9LUD7	Q9lud7 arabidopsis
43	38	46.3	227	2 Q9HV03	Q9hv03 pseudomonas
44	38	46.3	259	2 O86303	O86303 streptomyce
45	38	46.3	277	2 Q9ZBW7	Q9zbw7 streptomyce
46	38	46.3	371	2 Q9L8R0	Q9l8r0 pseudomonas
47	38	46.3	387	2 Q9X0N1	Q9x0n1 thermotoga
48	38	46.3	417	2 Q9XC39	Q9xc39 pseudomonas
49	38	46.3	419	5 O44549	O44549 caenorhabdi
50	38	46.3	460	14 Q65854	Q65854 beet yellow
51	38	46.3	490	5 O18169	O18169 caenorhabdi
52	38	46.3	561	1 Q9V2L7	Q9v2l7 pyrococcus
53	38	46.3	635	2 Q9X283	Q9x283 thermotoga
54	38	46.3	1049	2 O31100	O31100 pseudomonas
55	38	46.3	1408	10 Q9FWU0	Q9fwu0 oryza sativ
56	38	46.3	1930	10 Q9SHJ3	Q9shj3 arabidopsis
57	38	46.3	2140	3 Q02316	Q02316 lentinula e
58	38	46.3	2581	5 Q9N4C8	Q9n4c8 caenorhabdi
59	38	46.3	4564	5 O77075	O77075 drosophila
60	37	45.1	164	2 Q9K5R8	Q9k5r8 bacillus ha
61	37	45.1	196	2 Q9EVK0	Q9evk0 sinorhizobi
62	37	45.1	207	2 O69082	O69082 pseudomonas
63	37	45.1	225	1 O29453	O29453 archaeoglob
64	37	45.1	266	2 Q9X8J4	Q9x8j4 streptomyce
65	37	45.1	288	10 Q9ZQB0	Q9zqb0 arabidopsis

ALIGNMENTS

RESULT	1	
Q9VSV1	ID	Q9VSV1
AC	Q9VSV1;	PRELIMINARY; PRT; 599 AA.
DT	01-MAY-2000 (TREMBlrel. 13, Created)	
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)	
DT	01-JUN-2000 (TREMBlrel. 14, Last annotation update)	
DE	CG4484 PROTEIN.	
GN	CG4484.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BERKELEY;	
RX	MEDLINE-20196006; PubMed-10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Bortova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Zheng L.,  
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003552; AAF50310.1; -;  
DR FlyBase; FBgn0035968; CG4484.  
SQ SEQUENCE 599 AA; 66057 MW; C5381D334CFBF2EB CRC64;

Query Match 59.8%; Score 49; DB 5; Length 599;  
Best Local Similarity 50.0%; Pred. No. 5.1;  
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEKF 16  
Db 72 FVSPILLQIGVDHKH 87

RESULT 2 PRELIMINARY; PRT; 280 AA.  
ID Q9RKH8  
AC Q9RKH8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE PUTATIVE ABC-TRANSPORTER INNER MEMBRANE PROTEIN.  
GN SCE46.11c.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
[1] SEQUENCE FROM N.A.  
RP STRAIN-A3(2);  
RC Brown S.P., Harris D.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; ALI33252; CAB61862.1; -;  
DR InterPro; IPR000515; -;  
DR Pfam; PF0528; BPD\_TRANSF\_1; INN\_MEMBR; UNKNOWN\_1.  
DR PROSITE; PS00402; BPD\_TRANSF\_1; INN\_MEMBR; UNKNOWN\_1.  
SQ SEQUENCE 280 AA; 29074 MW; 3D00572E5DD83B56 CRC64;

Query Match 56.1%; Score 46; DB 2; Length 280;  
Best Local Similarity 60.0%; Pred. No. 7.4;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLLEVGVEEKF 16  
Db 108 VPPFALAVGVQEVFV 122

RESULT 3 PRELIMINARY; PRT; 299 AA.  
ID Q56171  
AC Q56171; Q56162;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE N-ACETYLPHOSPHINOETHRIN-TRIPETIDE- DEACETYLASE (PHOSPHINOETHRIN N-  
ACETYLTRANSFERASE) (PAT).  
GN DEA.  
OS Streptomyces viridochromogenes.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1938;  
[1] SEQUENCE FROM N.A.  
RP STRAIN-TU494;  
RC MEDLINE=96156798; PubMed=8593056;  
RX Schwartz D., Alijah R., Nussbaumer B., Pelzer S., Wohlleben W.;  
RA "The peptide synthetase gene phsA from Streptomyces viridochromogenes  
is not juxtaposed with other genes involved in nonribosomal  
biosynthesis of peptides.";  
RT Appl. Environ. Microbiol. 62:570-577(1996).  
RN [2]  
RP SEQUENCE OF 1-164 FROM N.A.  
RC STRAIN-TU494;  
RX MEDLINE=89196914; PubMed=3240868;  
RA Wohlleben W., Arnold W., Broer I., Hillemann D., Strauch E.,  
RA Puehler A.;  
RT "Nucleotide sequence of the phosphinoethrin N-acetyltransferase gene  
from Streptomyces viridochromogenes Tu494 and its expression in  
Nicotiana tabacum.";  
RL Gene 70:25-37(1988).  
DR EMBL; X65195; CAA46315.1; -;  
DR InterPro; IPR000379; -;  
DR InterPro; IPR002168; -;  
DR PROSITE; PS01173; LIPASE-GDXG\_HIS; 1.  
DR PROSITE; PS01174; LIPASE-GDXG\_SER; 1.  
SQ SEQUENCE 299 AA; 32203 MW; 4CC2D182E0882E22 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 299;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPPLLLEVGVEE 13  
Db 226 LPPLLIQVGSEE 237

RESULT 4  
Q9LSL1 ID Q9LSL1 PRELIMINARY; PRT; 351 AA.  
AC Q9LSL1;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE GENOMIC DNA, CHROMOSOME 5, TAC CLONE:K21L13.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB026639; BAA98183.1; -.  
DR InterPro; IPR000515; -.  
DR InterPro; IPR001092; -.  
DR InterPro; IPR003015; -.  
DR Pfam; PF00010; HLH; 1.  
DR PROSITE; PS00402; BPD\_TRANSF\_INN\_MEMBR; UNKNOWN\_1.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
DR SMART; SM00353; HLH; 1.  
SQ SEQUENCE 351 AA; 39484 MW; A28579226445A459 CRC64;

Query Match 52.4%; Score 43; DB 10; Length 351;  
Best Local Similarity 61.5%; Pred. No. 31;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLEVGVEEKF 15  
| | | : | | | |  
Db 124 PPLILQNGQEEF 136

RESULT 5  
Q9HLZ9 ID Q9HLZ9 PRELIMINARY; PRT; 230 AA.  
AC Q9HLZ9;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PROTEASE (PSPA) RELATED PROTEIN.  
GN TA0074.  
OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;  
OC Thermoplasma.  
OX NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 1728;  
RX MEDLINE=20479972; PubMed=11029001;  
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
RT acidophilum.";  
RL Nature 407:508-513(2000).  
DR EMBL; AL445063; CAC11222.1; -.  
DR InterPro; IPR002142; -.  
DR Pfam; PF01343; Peptidase\_U7; 1.  
SQ SEQUENCE 230 AA; 26584 MW; 6C60B3F66D952DA4 CRC64;

Query Match 51.2%; Score 42; DB 1; Length 230;  
Best Local Similarity 57.1%; Pred. No. 29;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEK 14  
| : | | | : | | | |  
Db 107 YIKPLLDKIGVEMK 120  
  
RESULT 6  
Q9FC55 ID Q9FC55 PRELIMINARY; PRT; 316 AA.  
AC Q9FC55;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PUTATIVE LIPASE.  
GN SC4B10.32.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Saunders D.C., Harris D.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL391515; CAC04237.1; -.  
SQ SEQUENCE 316 AA; 34103 MW; E9E20D2C8766549A CRC64;

Query Match 51.2%; Score 42; DB 2; Length 316;  
Best Local Similarity 80.0%; Pred. No. 41;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 PPLLEVGVE 12  
| | | | | : | | | |  
Db 20 PPLLYELGVE 29

RESULT 7  
Q9WYC6 ID Q9WYC6 PRELIMINARY; PRT; 367 AA.  
AC Q9WYC6;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE CITRATE SYNTHASE (EC 4.1.3.7).  
GN TM0290.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from



RT genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +  
CC OXALOACETATE.  
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
CC -1- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS  
CC CAPABLE OF OXIDATIVE METABOLISM.  
CC -1- SIMILARITY: TO THE CITRATE SYNTHASE FAMILY.  
CC EMBL: AE001711; AAD35378.1; --  
CC HSSP: Q53554; 1AJ8.  
CC TIGR: TW0290; --  
DR InterPro; IPR000023; --  
DR InterPro; IPR002020; --  
DR Pfam; PF00285; citrate\_synt; 1.  
DR PRINTS; PR00143; CITRITSNTHASE.  
DR PRODOM; PD000707; --; 1.  
DR PROSITE; PS00480; CITRATE SYNTHASE; 1.  
DR Lyase; Tricarboxylic acid cycle.  
KW SEQUENCE 367 AA; 42304 MW; C95CDAFC6D3D7D8F CRC64;

Query Match 51.2%; Score 42; DB 2; Length 367;  
Best Local Similarity 53.8%; Pred. No. 48;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEK 14  
Db 218 VPPMLEIGSEDR 230

RESULT 8  
Q9X3R1 PRELIMINARY; PRT; 449 AA.  
ID Q9X3R1  
AC Q9X3R1, 1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PUTATIVE HALOGENASE.  
GN PLTA.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PF-5;  
RX MEDLINE-98094250; PubMed-9434161;  
RA Nowak-Thompson B., Gould S.J., Loper J.E.;  
RT "Identification and sequence analysis of the genes encoding a  
RT polyketide synthase required for pyoluteorin biosynthesis in  
RT Pseudomonas fluorescens Pf-5.";  
RL Gene 204:17-24(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PF-5;  
RX MEDLINE-99194726; PubMed-10094695;  
RA Nowak-Thompson B., Chaney N., Wing J.S., Gould S.J., Loper J.E.;  
RT "Characterization of the pyoluteorin biosynthetic gene cluster of  
RT Pseudomonas fluorescens Pf-5.";  
RL J. Bacteriol. 181:2166-2174(1999).  
DR EMBL; AF081920; AAD24884.1; --  
DR InterPro; IPR002025; --  
DR InterPro; IPR002938; --  
DR InterPro; IPR003042; --  
DR Pfam; PF01494; FAD\_binding\_3; 1.  
DR PRINTS; PR00420; RINGMNOXGNASE.  
DR SEQUENCE 449 AA; 50676 MW; 0D761E587B9B16DE CRC64;

Query Match 51.2%; Score 42; DB 2; Length 449;  
Best Local Similarity 72.7%; Pred. No. 59;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEK 14  
Db 54 PVLLEIGVMEK 64  
RESULT 9  
O16518 PRELIMINARY; PRT; 528 AA.  
ID O16518  
AC O16518;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE T05H4.13 PROTEIN.  
DE T05H4.13  
GN Caenorhabditis elegans.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE-94150718; PubMed-7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favellio A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., R.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Smaison N., Smith A., Sonnenhammer E., Staden R., Sulston J.,  
RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Waterston R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF016452; AAB66022.1; --  
DR HSSP; P11883; 1AD3.  
DR InterPro; IPR002086; --  
DR Pfam; PF00171; aldedh; 1.  
DR SEQUENCE 528 AA; 59688 MW; 1BE0A683BA29A8F8 CRC64;

Query Match 51.2%; Score 42; DB 5; Length 528;  
Best Local Similarity 43.8%; Pred. No. 70;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKFM 16  
Db 316 YIPPTVLDVEXSDPFM 331

RESULT 10  
Q9I338 PRELIMINARY; PRT; 700 AA.  
ID Q9I338  
AC Q9I338; 2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL PROTEIN PA1689.  
GN PA1689.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.

```
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004596; AAG05078.1; -.
KW Hypothetical protein.
SQ SEQUENCE 700 AA; 77984 MW; 08506AA9CB310FB4 CRC64;

Query Match 51.2%; Score 42; DB 2; Length 700;
Best Local Similarity 57.1%; Pred. No. 95;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEKF 15
   |||||:|:|
Db 559 VPLLIIAPGIEKF 572

RESULT 11
Q54386 PRELIMINARY; PRT; 260 AA.
AC Q54386;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE LINCOMYCIN RESISTANCE METHYLASE.
GN LRM.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92104506; PubMed=1761231;
RA Jenkins G., Cundliffe E.;
RT "Cloning and characterization of two genes from Streptomyces lividans
RT that confer inducible resistance to lincomycin and macrolide
RT antibiotics.";
RL Gene 108:55-62(1991).
DR EMBL; M74717; AAA26779.1; -.
DR InterPro; IPR000051; -.
DR InterPro; IPR001737; -.
DR Pfam; PF00398; RrnaAD; 1.
DR PROSITE; PS01131; RRNA_A_DIMETH; 1.
KW Methyltransferase.
SQ SEQUENCE 260 AA; 28837 MW; 83345F6452DF2215 CRC64;

Query Match 50.0%; Score 41; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PPLLEVG 10
   |||||
Db 36 PPLLEVG 43

RESULT 12
Q9I217 PRELIMINARY; PRT; 309 AA.
AC Q9I217;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
```

```
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PROBABLE ESTERASE/DEACETYLASE.
GN PA2098.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004637; AAG05486.1; -.
DR InterPro; IPR000379; -.
DR InterPro; IPR002168; -.
DR PROSITE; PS01173; LIPASE_GDXG_HIS; 1.
DR PROSITE; PS01174; LIPASE_GDXG_SER; UNKNOWN_1.
SQ SEQUENCE 309 AA; 33018 MW; 69CAFF3CB0F9CD33 CRC64;
```

Query Match 50.0%; Score 41; DB 2; Length 309;  
Best Local Similarity 58.3%; Pred. No. 59;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEE 13  
:||||:|:|  
Db 232 LPPLLQVGEDE 243

```
RESULT 13
Q9X6R2 PRELIMINARY; PRT; 405 AA.
ID Q9X6R2
AC Q9X6R2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H103;
RA Adewoye L.O., Worobec E.A.;
RT "Identification of a trans-regulatory locus involved in the regulation
RT of carbohydrate transport in Pseudomonas aeruginosa.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF143948; AAD32694.1; -.
SQ SEQUENCE 405 AA; 44968 MW; BF441FA50E1B7D58 CRC64;
```

Query Match 50.0%; Score 41; DB 2; Length 405;  
Best Local Similarity 57.1%; Pred. No. 78;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEKF 15  
|||:|:|  
Db 264 VPLLIGPGIEKF 277

```
RESULT 14
O67814 PRELIMINARY; PRT; 436 AA.
ID O67814
AC O67814;
```

DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE PROTOPORPHYRINOGEN OXIDASE.  
CN HEMG.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
PC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus";  
RL Nature 392:353-358(1998).  
DR EMBL: AE000768; AAC07778.1; --  
DR InterPro; IPR002025; --  
DR InterPro; IPR002937; --  
DR Pfam; PF01593; Amino\_oxidase; 1.  
SQ SEQUENCE 436 AA; 48987 MW; 1. ED5F2B1BCD1DEEF7 CRC64;

Query Match 50.0%; Score 41; DB 2; Length 436;  
Best Local Similarity 53.3%; Pred. No. 85;  
Matches 8; Conservative 3; Mismatches 4; Indels 4; Gaps 0;

QY 1 YVPPLLEVGVEEKF 15  
| | : : | | | | |  
Db 279 YPPVVVVVNVGVEGKF 293

RESULT 15  
ID O74056 PRELIMINARY; PRT; 3472 AA.  
AC O74056;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 367.1 KDA PROTEIN.  
OS Cenarchaeum symbiosum.  
OC Archaea; Crenarchaeota; Cenarchaeum.  
OX NCBI\_TaxID=46770;  
RN [1]  
RP SEQUENCE FROM N.A.  
PC STRAIN=B;  
RX MEDLINE=98422450; PubMed=9748430;  
RA Schlieper C., Delong E.F., Preston C.M., Feldman R.A., Wu K.Y.,  
RA Swanson R.V.;  
RT "Genomic analysis reveals chromosomal variation in natural populations  
of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";  
RL J. Bacteriol. 180:5003-5009(1998).  
DR EMBL: AF083072; AAC62699.1; --  
DR InterPro; IPR000515; --  
DR InterPro; IPR001680; --  
DR Pfam; PF00400; WD40; 5.  
DR PROSITE; PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.  
DR SMART; SM00320; WD40; 1.  
KW Hypothetical protein; Repeat; WD repeat.  
SQ SEQUENCE 3472 AA; 367061 MW; 37F80707030F9355 CRC64;

Query Match 50.0%; Score 41; DB 1; Length 3472;  
Best Local Similarity 66.7%; Pred. No. 7.8e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEE 13  
| | | | | | | |  
Db 1169 LPPLFLGVGAEE 1180

RESULT 16  
ID Q08504 PRELIMINARY; PRT; 108 AA.  
AC Q08504;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE ORF YOR105W.  
GN YOR105W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Voss H., Benes V., Rechmann S., Teodoru C., Schwager C., Paces V.,  
RA Ansoirge W.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
PC MIPS;  
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z75012; CAA99302.1; --  
DR SGD; S0005631; YOR105W.  
SQ SEQUENCE 108 AA; 12335 MW; F89561B3720B66A5 CRC64;

Query Match 48.8%; Score 40; DB 3; Length 108;  
Best Local Similarity 40.0%; Pred. No. 28;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEKF 16  
| | | : | | : | :  
Db 69 LPPLIEVHIQPKII 83

RESULT 17  
ID Q56530 PRELIMINARY; PRT; 136 AA.  
AC Q56530;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE ALPHA SUBUNIT OF DINITROGENASE (MOFE PROTEIN) (FRAGMENT).  
GN NIFD.  
OS unidentified nitrogen-fixing bacteria.  
OC Bacteria.  
OX NCBI\_TaxID=34107;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ueda T.;  
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. PubMed=7736355;  
RX MEDLINE=95254398; Yahiro N., Matsuguchi T.;  
RA Ueda T., Suga Y., Yahiro N., Matsuguchi T.;  
RT "Genetic diversity of N2-fixing bacteria associated with rice roots by  
molecular evolutionary analysis of a nifD library.";  
RL Can. J. Microbiol. 41:235-240(1995).  
DR EMBL; D26279; BAA05341.1; --  
DR HSSP; P07328; 3MIN.  
DR InterPro; IPR000510; --  
DR Pfam; PF00148; oxidored\_nitro; 1.  
DR NON\_TER 1  
FT NON\_TER 136  
SQ SEQUENCE 136 AA; 15413 MW; E4AA82415AE9CF2A CRC64;

Query Match 48.8%; Score 40; DB 2; Length 136;  
Best Local Similarity 46.7%; Pred. No. 36;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEKF 16





DR EMBL: AP000064; BAA81602.1; --  
DR InterPro; IPR001279; --  
DR Pfam; PF00753; lactamase\_B; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 348 AA; 37793 MW; 22086DAD85ABDE15 CRC64;  
  
Query Match 48.8%; Score 40; DB 1; Length 348;  
Best Local Similarity 81.8%; Pred. No. 99;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 5 LLELVGVVEEKF 15  
| | | | | | | | | |  
Db 98 LLLRVGVVEEYF 108  
  
RESULT 22  
Q9N389 PRELIMINARY; PRT; 407 AA.  
ID Q9N389  
AC Q9N389;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE HYPOTHETICAL PROTEIN Y55B1A1.A.  
GN Y55B1A1.A.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AC024819; AAF59587.1; --  
SQ SEQUENCE 407 AA; 45724 MW; 0B5FF18982C2A848 CRC64;  
  
Query Match 48.8%; Score 40; DB 5; Length 407;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 5 LLELVGVVEEKF 16  
| | | | | | | | | |  
Db 219 LLAECGLEEKF 230  
  
RESULT 23  
Q96544 PRELIMINARY; PRT; 437 AA.  
ID Q96544  
AC Q96544;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE CITRATE (SI)-SYNTHASE (EC 4.1.3.7) (CONDENSING ENZYME) (CITRATE  
DE CONDENSING ENZYME) (CITROGENASE) (OXALOACETATE TRANSACETASE)  
DE (FRAGMENT).  
GN Cit1.  
OS Beta vulgaris (Sugar beet).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;  
OC Caryophyllales; Chenopodiaceae; Beta.  
OX NCBI\_TaxID=3555;  
RN [1]  
RP SEQUENCE FROM N.A.

RA La Cognata U., Landschuetze V., Willmitzer L., Mueller-Roeber B.;  
RL Plant Cell Physiol. 0:0-0(0).  
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +  
CC OXALOACETATE.  
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE (BY SIMILARITY).  
CC -1- SIMILARITY: TO THE CITRATE SYNTHASE FAMILY.  
DR EMBL: X84228; CAA59010.1; --  
DR HSSP; P23007; 2CSC.  
DR Mendel; 15201; Betvu; 2977; 15201.  
DR InterPro; IPR002020; --  
DR Pfam; PF00285; citrate\_synt; 1.  
DR PRINTS; PR00143; CITRITSNTHASE.  
KW Lyase; Tricarboxylic acid cycle.  
FT NON\_TER 1  
SQ SEQUENCE 437 AA; 48881 MW; FB50CA21ED5519A3 CRC64;  
  
Query Match 48.8%; Score 40; DB 10; Length 437;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 VPPLLEVG 10  
| | | | | | | | | |  
Db 356 VPPILLELG 364  
  
RESULT 24  
Q9JX05 PRELIMINARY; PRT; 687 AA.  
ID Q9JX05  
AC Q9JX05;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE GLCYL-TRNA SYNTHETASE, BETA CHAIN.  
GN NMB1930.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / SEROGROUP B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Pedersen J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,  
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain  
MC58.";  
RT Science 287:1809-1815(2000).  
RL EMBL: AE002541; AAF42259.1; --  
DR TIGR; NMB1930; --  
DR InterPro; IPR002311; --  
DR Pfam; PF02092; tRNA\_synt\_2f; 1.  
DR PRINTS; PR01045; TRNASYNTHGB.  
KW Aminoacyl-tRNA synthetase.  
SQ SEQUENCE 687 AA; 74573 MW; 616BCBDD76A3D4FB CRC64;  
  
Query Match 48.8%; Score 40; DB 2; Length 687;  
Best Local Similarity 53.8%; Pred. No. 2e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 PLLLEVGVEEKF 16  
| | | | | | | | | |  
Db 259 PVLLEAGFEHFL 271  
  
RESULT 25  
Q9JW67

ID Q9JW67 PRELIMINARY; PRT; 687 AA.  
AC Q9JW67;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE PUTATIVE GLYCYL-TRNA SYNTHETASE BETA SUBUNIT (EC 6.1.1.14).  
GN GLYS OR NMA0523.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;  
RX MEDLINE=20222556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrell B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
meningitidis Z2491.";  
RL Nature 404:502-506(2000).  
DR EMBL; AL162753; CAB83815.1; -.  
DR InterPro; IPR002311; -.  
DR Pfam; PF02092; tRNA\_synt\_2f; 1.  
DR PRINTS; PR01045; TRNASYNTHGB.  
KW Aminoacyl-tRNA synthetase; Ligase.  
SQ SEQUENCE 687 AA; 74650 MW; 8C17811D5C93EBE3 CRC64;

Query Match 48.8%; Score 40; DB 2; Length 687;  
Best Local Similarity 53.8%; Pred. No. 2e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVVEKFM 16  
|:|:| | | | | |  
Db 259 PVLVLEAGFEEHFL 271  
  
RESULT 26  
O51274 PRELIMINARY; PRT; 737 AA.  
ID O51274  
AC O51274;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE HYPOTHETICAL 86.3 KDA PROTEIN.  
GN BB0259.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35210 / B31;  
RX MEDLINE=98065943; PubMed=9403685;  
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
burgdorferi.";  
RL Nature 390:580-586(1997).  
DR EMBL; AE001136; AAC66641.1; -.  
DR TIGR; BB0259; -.  
DR InterPro; IPR000189; -.  
DR Pfam; PF01464; SLT; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 737 AA; 86336 MW; 8D0D66B4F6A928A4 CRC64;

Query Match 48.8%; Score 40; DB 2; Length 737;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLLLEVGVVEKFM 16  
| | | | : | | | | |  
Db 94 YVAYLFFKKIGFEKFM 109

RESULT 27  
Q9LCT3 PRELIMINARY; PRT; 916 AA.  
ID Q9LCT3  
AC Q9LCT3;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE PREPROTEIN TRANSLOCASE (SECRETION PROTEIN SECA).  
GN SECA OR PA4403.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=96146052; PubMed=8581173;  
RA Liao X., Charlebois I., Ouellet C., Morency M.J., Dewar K.,  
RA Lightfoot J., Foster J., Siehnel R., Schweizer H., Lam J.S.,  
RA Hancock R.E., Levesque R.C.;  
RT "Physical mapping of 32 genetic markers on the Pseudomonas aeruginosa  
PA01 chromosome.";  
RT Microbiology 142:79-86(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RA Levesque R.C., Liao X., Lightfoot J., Charlebois I., Ouellet C.,  
RA Morency M., Dewar K., Siehnel R., Lam J., Hancock R.E.;  
RT "Physical Mapping of 38 loci including aimE, ampC, ampR, arcA, aroK,  
catR, cypH, dapB, envA, envC, ftsA, ftsZ, groEL, mreE, oprD,  
oprF, oprH, oprI, oprK, oprP, pbpB, pbpC, pheS, phoA, phoB, phoS,  
ponA, pyoSl, qin, rpoB, rpoH, sodB, soxR, soxC.";  
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; U19797; AAF26459.1; -.  
DR EMBL; AE004856; AAG07791.1; -.  
DR InterPro; IPR000185; -.  
DR InterPro; IPR001410; -.  
DR Pfam; PF01043; SecA\_protein; 1.  
DR PRINTS; PR00906; SECA.  
DR PROSITE; PS01312; SECA; 1.  
SQ SEQUENCE 916 AA; 103854 MW; D2B2FE82D07F651F CRC64;

Query Match 48.8%; Score 40; DB 2; Length 916;  
Best Local Similarity 50.0%; Pred. No. 2.8e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 YVPPLLLLEVGVVEKFM 16  
| | | | | | | | | |

DE 464 YVSKLLQEGIEHKVL 479  
 DE 28  
 Q9K5S0  
 ID Q9K5S0 PRELIMINARY; PRT; 162 AA.  
 AC Q9K5S0  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE BH4018 PROTEIN.  
 GN BH4018.  
 OS Bacillus halodurans.  
 OC Bacillus; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RA Takami H., Nakasone K., Takaki Y.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP001520; BAB07737.1; -. 119BC6D425F4118B CRC64;  
 SQ SEQUENCE 162 AA; 18884 MW; 119BC6D425F4118B CRC64;

Query Match 47.6%; Score 39; DB 2; Length 162;  
 Best Local Similarity 54.5%; Pred. No. 65;  
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEK 14  
 Db 151 PVLIELGLEQK 161

RESULT 29  
 ID Q9U0L8 PRELIMINARY; PRT; 286 AA.  
 AC Q9U0L8  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE HYPOTHETICAL 33.9 KDA PROTEIN.  
 GN MAL4P2.04.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,  
 RA Quail M., Barrell B.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035475; CAB62845.2; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 286 AA; 33896 MW; 7901FB800BFC485B CRC64;

Query Match 47.6%; Score 39; DB 5; Length 286;  
 Best Local Similarity 35.7%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEK 14  
 Db 76 YIPAIWKYGIEER 89

RESULT 30  
 ID Q95685 PRELIMINARY; PRT; 299 AA.  
 AC Q95685  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)

SERINE-THREONINE SPECIFIC PROTEIN PHOSPHATASE (EC 3.1.3.16)  
 (DJ551D2.3) (PROTEIN PHOSPHATASE 1, REGULATORY SUBUNIT 6).  
 PPPIR6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98074939; PubMed=9414128;  
 RA Armstrong C.G., Browne G.J., Cohen P., Cohen P.T.W.;  
 RL "PPPIR6, a novel member of the family of glycogen-targeting subunits  
 of protein phosphatase 1";  
 FEBS Lett. 418:210-214(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Lovell J.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y18206; CAA77081.1; -.  
 DR EMBL; AL109928; CAB92096.1; -.  
 KW Hydrolyase. 299 AA; 32559 MW; DB848FBLCF55E49E CRC64;  
 SQ SEQUENCE 299 AA; 32559 MW; DB848FBLCF55E49E CRC64;

Query Match 47.6%; Score 39; DB 4; Length 299;  
 Best Local Similarity 57.1%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEK 15  
 Db 242 VPPELLELGSRVHF 255

RESULT 31  
 ID P72208 PRELIMINARY; PRT; 307 AA.  
 AC P72208  
 DT 01-FEB-1997 (Tremblrel. 02, Created)  
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE ASPARTOACYLASE.  
 GN ASPA.  
 OS Prochlorococcus marinus.  
 OC Bacteria; Cyanobacteria; prochlorophytes; Prochlorococcaceae;  
 OC Prochlorococcus.  
 OX NCBI\_TaxID=1219;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CCMP 1375;  
 RX MEDLINE=97399394; PubMed=9255521;  
 RA Hess W.R.;  
 RL "Localization of an open reading frame with homology to human  
 aspartoacylase upstream from psba in the prokaryote Prochlorococcus  
 marinus CCMP 1375";  
 RL DNA Seq. 7:301-306(1997).  
 DR EMBL; Z80110; CAB02200.1; -.  
 SQ SEQUENCE 307 AA; 34213 MW; 4711407677490957 CRC64;

Query Match 47.6%; Score 39; DB 2; Length 307;  
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEEK 16  
 Db 256 LLEGGLEEFV 267

RESULT 32  
 ID Q9KXR2 PRELIMINARY; PRT; 326 AA.  
 AC Q9KXR2  
 DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE ASPARTATE CARBAMOYLTRANSFERASE.  
GN PYRB.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Brown S.P., Harris D.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL357523; CAB93367.1; -.  
DR InterPro; IPR002029; -.  
DR PRINTS; PRO0100; AOTCASE.  
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; UNKNOWN\_1.  
KW Transferase.  
SQ SEQUENCE 326 AA; 35661 MW; F353C20B27B43AB0 CRC64;  
  
Query Match 47.6%; Score 39; DB 2; Length 326;  
Best Local Similarity 80.0%; Pred. No. 1.4e+02;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 3 PPLLEVGVE 12  
Db 197 PPTLLPVGVE 206  
  
RESULT 33  
Q9P727 PRELIMINARY; PRT; 333 AA.  
AC Q9P727;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PROBABLE ATP-SPECIFIC SUCCINYL-COA SYNTHETASE ALPHA SUBUNIT.  
GN 8D4.130.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: TO OTHER SUBUNITS (ALPHA) OF SUCCINYL-COA SYNTHETASE,  
CC OF MALATE-COA LIGASE AND TO ATP CITRATE-LYASE.  
DR EMBL; AL353819; CAB88555.1; -.  
DR InterPro; IPR000303; -.  
DR Pfam; PF00549; ligase-CoA; 1.  
DR PROSITE; PS01216; SUCCINYL\_COA\_LIG\_1; 1.  
DR PROSITE; PS00399; SUCCINYL\_COA\_LIG\_2; 1.

KW Ligase; Phosphorylation.  
SQ SEQUENCE 333 AA; 34738 MW; BA6FBF63AD918BDE CRC64;  
  
Query Match 47.6%; Score 39; DB 3; Length 333;  
Best Local Similarity 53.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 YVPPLLLLEVGVEE 13  
Db 106 FVPPLAAGIEE 118  
  
RESULT 34  
P91430 PRELIMINARY; PRT; 419 AA.  
ID P91430;  
AC P91430;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE SIMILAR TO MOLYBDOTERIN BIOSYNTHESIS MOEB PROTEINS.  
GN T03Fl.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Du Z., Le T.T.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U88169; AAB42231.1; -.  
DR InterPro; IPR000205; -.  
DR InterPro; IPR000594; -.  
DR InterPro; IPR002162; -.  
DR Pfam; PF00899; Thif\_family; 1.  
DR PROSITE; PS00065; D\_2\_HYDROXYACID\_DH\_1; UNKNOWN\_1.  
SQ SEQUENCE 419 AA; 46511 MW; BA61B08DB8AC8F4C CRC64;  
  
Query Match 47.6%; Score 39; DB 5; Length 419;  
Best Local Similarity 46.2%; Pred. No. 1.8e+02;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 VPPLLLLEVGVEEK 14  
Db 240 VPPLVVASGIDER 252  
  
RESULT 35  
Q21603



```

ID Q21603 PRELIMINARY; PRT; 531 AA.
AC Q21603;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE M88.1.1 PROTEIN.
GN Caenorhabditis elegans.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCHI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RP Substion J.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Connell M., Copsey T., Cooper J., Coulson A.,
RA Bonfield J., Burton S., Du Z., Durbin R., Favell A., Fulton L.,
RA Craxton M., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Gardner A., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Jones M., Kershaw J., McMurray A., Mortimore B., O'Callaghan M.,
RA Lightning J., Lloyd C., McMuray A., Roopra A., Saunders D., Showkeen R.,
RA Parsons J., Percy C., Rifken L., Roopra A., Staden R., Sulston J.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Whildman P.; III of C.
RA "2.2 Mb of contiguous nucleotide sequence from chromosome
RT elegans.";
RT Nature 368:32-38(1994).
RL EMBL; 234802; CAA84336.1;
DR InterPro; IPR002213;
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; UNKNOWN_1.
DR SEQUENCE 531 AA; 60249 MW; 156994087344D36E CRC64;
SQ
Query Match 47.6%; Score 39; DB 5; Length 531;
Best Local Similarity 53.3%; Pred. No. 2.3e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 YVPPLLLLEVGVEEKF 15.
DB 187 YVPPLMMESDDMGF 201
PRELIMINARY; PRT; 570 AA.
ID Q9XEE4
AC Q9XEE4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL SER-THR PROTEIN KINASE.
GN Arabidopsis thaliana (Mouse-ear cross).
OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCHI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99225673; PubMed=10207155;
RX Wang M.L., Belmonte S., Kim U., Dolan M., Morris J.W., Goodman H.M.;
RA "A cluster of ABA-regulated genes on Arabidopsis thaliana BAC
RT T07M07.";
RT Genome Res. 9:325-333(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.; White O., Eisen J.A.,
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF085279; AAD25928.1;
DR EMBL; AC018721; AAF18726.1;
DR HSSP; P27703; 2ERK.
DR InterPro; IPR000719;
DR InterPro; IPR002290;
DR Pfam; PF00069; pkinase; 2.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR SMART; SM00220; S_TKc; 1.
DR ATP-binding; Hypothetical protein; Kinase;
KW Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 570 AA; 64915 MW; CB33D89CFF428E2E CRC64;
QY 4 PLLLEVGVEEK 14
DB 188 PELLIDIGLEDK 198
PRELIMINARY; PRT; 606 AA.
ID Q9PLA4
AC Q9PLA4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE (ISOMERIZING).
GN Chlamydia muridarum.
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCHI_TaxID=83560;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935; Gill S.R., Heidelberg J.F., Bass S.,
RA Read T.D., Brunham R.C., Shen C., Peterson J., Utterback T., Berry S.,
RA White O., Hickey E.K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Linher K., Weidman J., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Gwinn M., Nelson W., Fraser C.M.;
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL pneumoniae AK39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002287; AAF39075.1;
DR TIGR; TC0203;
DR InterPro; IPR000583;
DR InterPro; IPR001347;
DR Pfam; PF00310; GATase_2; 1.
DR Pfam; PF01380; SIS; 2.
DR PROSITE; PS00443; GATASE_TYPE_II; UNKNOWN_1.
DR SEQUENCE 606 AA; 67568 MW; 2B6469726717C2AC CRC64;
SQ
Query Match 47.6%; Score 39; DB 2; Length 606;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 YVPPLLLLEVGVEEKF 16
DB 175 YESPLILGLGEEVEFI 190
PRELIMINARY; PRT; 606 AA.
ID Q9PLA4
AC Q9PLA4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE (ISOMERIZING).
GN Chlamydia muridarum.
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCHI_TaxID=83560;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935; Gill S.R., Heidelberg J.F., Bass S.,
RA Read T.D., Brunham R.C., Shen C., Peterson J., Utterback T., Berry S.,
RA White O., Hickey E.K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Linher K., Weidman J., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Gwinn M., Nelson W., Fraser C.M.;
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL pneumoniae AK39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002287; AAF39075.1;
DR TIGR; TC0203;
DR InterPro; IPR000583;
DR InterPro; IPR001347;
DR Pfam; PF00310; GATase_2; 1.
DR Pfam; PF01380; SIS; 2.
DR PROSITE; PS00443; GATASE_TYPE_II; UNKNOWN_1.
DR SEQUENCE 606 AA; 67568 MW; 2B6469726717C2AC CRC64;
SQ
Query Match 47.6%; Score 39; DB 2; Length 606;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 YVPPLLLLEVGVEEKF 16
DB 175 YESPLILGLGEEVEFI 190
```

RESULT 38  
Q9H6U0  
ID Q9H6U0 PRELIMINARY; PRT; 674 AA.  
AC Q9H6U0;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CDNA: FLJ21877 FIS, CLONE HEP02695.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK025530; BAB15162.1; -.  
SQ SEQUENCE 674 AA; 75965 MW; E1AFFD757D82C1AA CRC64;

Query Match 47.6%; Score 39; DB 4; Length 674;  
Best Local Similarity 60.0%; Pred. No. 3e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;  
QY 2 VPPLLEVGVEEKF 16  
||||| :| |||:  
Db 301 VPPLLCNMG--EKFL 313

RESULT 39  
Q9XWL9  
ID Q9XWL9 PRELIMINARY; PRT; 800 AA.  
AC Q9XWL9;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE Y38F1A.3 PROTEIN.  
GN Y38F1A.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wallis J.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL; AL032639; CAA21636.1; -.  
DR InterPro; IPR000731; -.  
DR InterPro; IPR003392; -.  
DR Pfam; PF02460; Patched; 1.  
SQ SEQUENCE 800 AA; 90816 MW; 738E16F28DE42849 CRC64;

Query Match 47.6%; Score 39; DB 5; Length 800;  
Best Local Similarity 46.7%; Pred. No. 3.6e+02;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
QY 2 VPPLLEVGVEEKF 16  
||||| :| |||:  
Db 232 VPFLILSIGVDDVFI 246  
RESULT 40  
Q9H6W7  
ID Q9H6W7 PRELIMINARY; PRT; 1148 AA.  
AC Q9H6W7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CDNA: FLJ21792 FIS, CLONE HEP00441.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK025445; BAB15133.1; -.  
SQ SEQUENCE 1148 AA; 128549 MW; 5F6E1928D0BDEB7C CRC64;

Query Match 47.6%; Score 39; DB 4; Length 1148;  
Best Local Similarity 60.0%; Pred. No. 5.2e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;  
QY 2 VPPLLEVGVEEKF 16  
||||| :| |||:  
Db 775 VPPLLCNMG--EKFL 787

RESULT 41  
Q9LXT9  
ID Q9LXT9 PRELIMINARY; PRT; 1963 AA.  
AC Q9LXT9;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CALLOSE SYNTHASE CATALYTIC SUBUNIT-LIKE PROTEIN.  
GN T24H18\_170.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Robben J., Grymonprez B., Volckaert G., Bancroft I.,  
RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL353013; CAB88264.1; -.  
DR InterPro; IPR003440; -.  
DR Pfam; PF02364; Glucan\_synthase; 1.  
SQ SEQUENCE 1963 AA; 226393 MW; 242F166954B90C51 CRC64;  
Query Match 47.6%; Score 39; DB 10; Length 1963;  
Best Local Similarity 50.0%; Pred. No. 9.3e+02;

Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 4 PLLLEGVVEKPF 15  
|:|:|:|:|:|  
Db 1628 PMLMEIGLERGF 1639

RESULT 42

Q9LUD7 ID Q9LUD7 PRELIMINARY; PRT; 1972 AA.

AC 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE GLUCAN SYNTHASE-LIKE PROTEIN.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;

RA Nakamura Y.; analysis of Arabidopsis thaliana chromosome 3. I. Sequence

RT features of the regions of 4,504,864 bp covered by sixty pl and TAC

RT clones.;

RL DNA Res. 7:131-135(2000).

DR EMBL; AB023038; BAB02389.1;

DR InterPro; IPR003440;

DR Pfam; PF02364; Glucan\_synthase; 1.

SQ SEQUENCE 1972 AA; 228012 MW; F30FB9C840A3F43B CRC64;

Query Match 47.6%; Score 39; DB 10; Length 1972;

Best Local Similarity 38.5%; Pred. No. 9.3e+02;

Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 4 PLLLEGVVEKPF 16  
|:|:|:|:|:|  
Db 1607 PMVMEIGLEKGF 1619

RESULT 43

Q9HV03 ID Q9HV03 PRELIMINARY; PRT; 227 AA.

AC 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE PROBABLE TRANSCRIPTIONAL REGULATOR.

GN PA4806.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PA01; PubMed-10984043;

RX MEDLINE-20437337; Erwin A.L., Mizoguchi S.D., Warrenner P.,

RA Stover C.K., Pham X.-Q.T., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Hickey M.J., Brinkman F.S.L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Garber R.L., Goltzy L., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Brody L.L., Coulter S.N., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Smith K.A., Spencer D.H., Hancock R.E.W., Lory S., Olson M.V.,

RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,

RT \*Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT Opportunistic pathogen.;

RL Nature 406:959-964(2000).

DR EMBL; AE004893; AAG08192.1;

DR InterPro; IPR000792;

DR Pfam; PF00196; GerE; 1.

DR PRINTS; PRO0038; HTHLUXR.

SQ SEQUENCE 227 AA; 25264 MW; 8043DB0923E969A9 CRC64;

Query Match 46.3%; Score 38; DB 2; Length 227;

Best Local Similarity 46.7%; Pred. No. 1.4e+02;

Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 YVPPLLEGVVEKPF 15  
:|:|:|:|:|  
Db 9 HAPTLLEYGIAERF 23

RESULT 44

O86303 ID O86303 PRELIMINARY; PRT; 259 AA.

AC 086303;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE MONO-METHYL TRANSFERASE (EC 2.1.1.48).

GN SRMA.

OS Streptomyces ambofaciens.

OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;

OC Actinomycetales; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1889;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC23877;

RA Gourmelon A., Blondelet-Rouault M.H., Guerin M., Pernodet J.L.;

RT "Characterization of a macrolide glycosyl transferase gene in

RT Streptomyces ambofaciens, the producer of the macrolide antibiotic

RT spiramycin.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

DR HSSP; AJ223970; CAA11706.1;

DR HSSP; P81989; IQE5.

DR InterPro; IPR000051;

DR InterPro; IPR001737;

DR Pfam; PF00398; RnaA; 1.

DR PROSITE; PS01131; RNA\_A\_DIMETH; 1.

KW Transferase; Methyltransferase.

SQ SEQUENCE 259 AA; 28637 MW; B0E61215F73E77E1 CRC64;

Query Match 46.3%; Score 38; DB 2; Length 259;

Best Local Similarity 87.5%; Pred. No. 1.6e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 PPLLEVG 10  
|:|:|:|:|  
Db 36 PPLVLEVG 43

RESULT 45

Q9ZBW7 ID Q9ZBW7 PRELIMINARY; PRT; 277 AA.

AC 09ZBW7;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE PUTATIVE OXIDOREDUCTASE.

GN SC4B5.01C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;

OC Actinomycetales; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
DR EMBL; AL034443; CAA22355.1; -.  
DR HSSP; P15121; 2ACQ.  
DR InterPro; IPR001395; -.  
DR Pfam; PF00248; aldo\_ket\_red; 1.  
DR PRINTS; PR00069; ALDKETREDTASE.  
DR PROSITE; PS00062; ALDOKETO\_REDUCTASE\_2; UNKNOWN\_1.  
DR PROSITE; PS00063; ALDOKETO\_REDUCTASE\_3; 1.  
SQ SEQUENCE 277 AA; 30227 MW; EE7A3D39474B60FE CRC64;

Query Match 46.3%; Score 38; DB 2; Length 277;  
Best Local Similarity 63.6%; Pred. No. 1.7e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12  
|||:| |||  
Db 5 VPPILNNGVE 15

Search completed: June 28, 2001, 11:55:30  
Job time: 151 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:38 ; Search time 72.61 Seconds  
(without alignments)  
18.221 Million cell updates/sec

Title: US-09-439-313-573  
Perfect score: 45  
Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SPTREMBL\_16:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	34	75.6	597	3	Q9HFG9	Q9hfg9 penicillium
2	33	73.3	3413	2	O54593	O54593 amycolatops
3	32	71.1	408	5	O43978	O43978 babesia bov
4	32	71.1	3469	5	Q9U4I2	Q9u4i2 drosophila
5	32	71.1	3502	5	Q9VYJ9	Q9vyj9 drosophila
6	31	68.9	123	4	Q9Y3N1	Q9y3n1 homo sapien
7	31	68.9	252	5	Q9VCV0	Q9vcv0 drosophila
8	31	68.9	394	5	O45782	O45782 caenorhabdi
9	31	68.9	436	2	Q9I4F5	Q9i4f5 pseudomonas
10	31	68.9	468	6	Q9N135	Q9n135 sus scrofa
11	31	68.9	470	6	Q9N129	Q9n129 bos taurus
12	31	68.9	500	2	Q9KWA5	Q9kwa5 agrobacteri
13	31	68.9	617	4	Q9NSZ9	Q9nsz9 homo sapien
14	30	66.7	68	14	Q9QJC6	Q9qjc6 human immun
15	30	66.7	148	5	Q9NM69	Q9nm69 leishmania
16	30	66.7	161	2	Q9RSZ9	Q9rsz9 deinococcus
17	30	66.7	170	8	O63327	O63327 bufo andrew
18	30	66.7	170	8	O63329	O63329 bufo andrew
19	30	66.7	170	8	O63331	O63331 bufo gargar

20	30	66.7	170	8	O63332	O63332 bufo gargar
21	30	66.7	170	8	O63333	O63333 bufo gargar
22	30	66.7	170	8	O63957	O63957 bufo gargar
23	30	66.7	170	8	O64012	O64012 bufo andrew
24	30	66.7	193	8	O21236	O21236 reclinomona
25	30	66.7	241	2	Q9HWT0	Q9hwt0 pseudomonas
26	30	66.7	261	2	Q9KJP3	Q9kjp3 myxococcus
27	30	66.7	269	14	Q9IPS9	Q9ips9 tobacco nec
28	30	66.7	277	14	O88611	O88611 tobacco nec
29	30	66.7	280	2	Q9PJ60	Q9pj60 campylobact
30	30	66.7	327	10	O48641	O48641 arabidopsis
31	30	66.7	364	3	Q9P367	Q9p367 neurospora
32	30	66.7	425	10	Q9SV55	Q9sv55 arabidopsis
33	30	66.7	473	2	Q9RK75	Q9rk75 streptomyce
34	30	66.7	503	2	Q9PDU7	Q9pdu7 xylella fas
35	30	66.7	539	1	Q9VD39	Q9vd39 aeropyrum p
36	30	66.7	587	5	Q9VGP3	Q9vgp3 drosophila
37	30	66.7	624	5	Q94486	Q94486 dictyosteli
38	30	66.7	662	4	Q9UFJ0	Q9ufj0 homo sapien
39	30	66.7	703	4	Q9NQF8	Q9nqf8 homo sapien
40	30	66.7	750	2	P74111	P74111 synechocyst
41	30	66.7	816	5	Q27537	Q27537 caenorhabdi
42	30	66.7	944	5	Q9VKQ6	Q9vkq6 drosophila
43	30	66.7	962	10	Q9LJX4	Q9ljx4 arabidopsis
44	30	66.7	964	10	Q9ZW02	Q9zw02 arabidopsis
45	30	66.7	966	10	Q9M6S0	Q9m6s0 populus tre
46	30	66.7	968	10	Q9ZW07	Q9zw07 arabidopsis
47	30	66.7	969	4	Q9Y2I4	Q9y2i4 homo sapien
48	30	66.7	972	10	Q9ZW06	Q9zw06 arabidopsis
49	30	66.7	1036	10	Q9FTP5	Q9ftp5 oryza sativ
50	30	66.7	1341	4	Q9UFP8	Q9ufp8 homo sapien
51	30	66.7	1671	5	Q9NUJ5	Q9njv5 biomphalari
52	30	66.7	2308	5	Q9VDJ9	Q9vdj9 drosophila
53	30	66.7	2559	5	O44113	O44113 drosophila
54	30	66.7	2559	5	O44381	O44381 drosophila
55	30	66.7	2559	5	Q9V7X3	Q9v7x3 drosophila
56	30	66.7	5060	2	O52545	O52545 amycolatops
57	30	66.7	5069	2	O52789	O52789 amycolatops
58	29	64.4	115	2	Q9ZFD5	Q9zfd5 riftia pach
59	29	64.4	125	3	Q9HE79	Q9he79 neurospora
60	29	64.4	126	4	Q9H7W1	Q9h7w1 homo sapien
61	29	64.4	160	2	Q9K5G3	Q9k5g3 alcaligenes
62	29	64.4	161	2	Q9K5G2	Q9k5g2 alcaligenes
63	29	64.4	188	2	Q59802	Q59802 staphylococ
64	29	64.4	210	2	Q9ZEN6	Q9zen6 wolinnella s
65	29	64.4	212	5	Q17223	Q17223 bombyx mori

ALIGNMENTS

RESULT 1					
Q9HFG9	ID	Q9HFG9	PRELIMINARY;	PRT;	597 AA.
AC	Q9HFG9;				
DT	01-MAR-2001	(TREMBLrel. 16, Created)			
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)			
DE	PUTATIVE ALPHA GLUCOSIDASE.				
GN	DEXC.				
OS	Penicillium minioluteum.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.				
OX	NCBI_TaxID=28574;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Garcia B., Castellanos A., Menendez J.;				
RT	"Isolation of a gene encoding putative alpha glucosidase from				
RT	Penicillium minioluteum."				
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ278706; CAC09327.1; -.				
SQ	SEQUENCE 597 AA; 68120 MW; 3F158E6544EA70C4 CRC64;				

```
Query Match      75.6%; Score 34; DB 3; Length 597;
Best Local Similarity 70.0%; Pred. NO. 28;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 1 MGSISEISQS 10

RESULT 2
O54593 PRELIMINARY; PRT; 3413 AA.
AC O54593;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE POLYKETIDE SYNTHASE.
GN RIFE.
OS Amycolatopsis mediterranei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX NCBI_TaxID=33910;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=S699;
RC Kim C.G., Yu T.W., Fryhle C., Handa S., Floss H.G.;
RA J. Biol. Chem. 0:0-0(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=S699;
RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,
RA Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,
RA Floss H.G.;
RN Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RN Chem. Biol. 5:0-0(0002).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=LBG A3136;
RA Schupp T., Toupet C., Engel N., Goff S.;
RA Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF040570; AAC01714.1;
DR EMBL; AJ223012; CAAL1039.1;
DR HSSP; P25715; IMLA.
DR InterPro; IPR000255;
DR InterPro; IPR000794;
DR InterPro; IPR001227;
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF00550; pp-binding; 2.
DR Pfam; PF00698; Acyl_transf; 2.
DR PROSITE; PS00075; ACP_DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_2.
KW Phosphopantetheine; Transferrase.
SQ SEQUENCE 3413 AA; 353729 MW; FFA25AC4A3920AAA CRC64;

Query Match      73.3%; Score 33; DB 2; Length 3413;
Best Local Similarity 60.0%; Pred. No. 3e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 2588 LGAVVQLPQS 2597

RESULT 3
O43978 PRELIMINARY; PRT; 408 AA.
AC O43978;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE DNAJ HOMOLOG.
OS Babesia bovis.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
OX NCBI_TaxID=5865;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MEXICO MO7;
RC Suarez C.E., Palmer G.H., McElwain T.F.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF027149; AAC27389.1;
DR HSSP; P25685; IHDJ.
DR InterPro; IPR001305;
DR InterPro; IPR001623;
DR InterPro; IPR002939;
DR InterPro; IPR003095;
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF00684; DnaJ_C; 1.
DR PRINTS; PR00625; DNAJPROTEIN.
DR PROSITE; PS00636; DNAJ_1; 1.
DR PROSITE; PS00076; DNAJ_2; 1.
DR SMART; SM00271; DnaJ; 1.
SQ SEQUENCE 408 AA; 45819 MW; 55DCFC389E0CFD6C CRC64;

Query Match      71.1%; Score 32; DB 5; Length 408;
Best Local Similarity 60.0%; Pred. NO. 54;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 169 MGAMIQQSQS 178

RESULT 4
O9U4I2 PRELIMINARY; PRT; 3469 AA.
AC O9U4I2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE SANT DOMAIN PROTEIN SMRTER.
GN SMR OR SMRTER OR CG4013.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RC MEDLINE=99417957; PubMed=10488333;
RA Tsai C.-C., Kao H.-Y., Yao T.-P., McKeown M., Evans R.M.;
RA "SMRTER, a Drosophila nuclear receptor coregulator, reveals that ECR-
RT mediated repression is critical for development.";
RL Mol. Cell 4:175-186(1999).
DR EMBL; AF175223; AAD52614.1;
DR FlyBase; FBgn0024308; Smr.
DR InterPro; IPR000104;
DR InterPro; IPR001005;
DR InterPro; IPR002086;
DR Pfam; PF00249; myb_DNA-binding; 1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
DR SMART; SM00395; SANT; 1.
SQ SEQUENCE 3469 AA; 364115 MW; 6284E14C5C247CD9 CRC64;

Query Match      71.1%; Score 32; DB 5; Length 3469;
```

Best Local Similarity 66.7%; Pred. No. 5.1e+02; Mismatches 2; Indels 0; Gaps 0; Matches 6; Conservative 1;

QY 2 GSIVQLSQS 10  
I :||:||||  
Db 291 GGWVQVSQS 299

RESULT 5  
Q9VYJ9 PRELIMINARY; PRT; 3502 AA.  
AC Q9VYJ9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE SMR PROTEIN.  
GN SMR OR CG4013.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskera D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003490; AAF48196.1; -.  
DR FlyBase; FBgn0024308; Smr.  
DR InterPro; IPR000104; -.  
DR InterPro; IPR001005; -.  
DR InterPro; IPR002086; -.  
DR Pfam; PF00249; myb\_DNA-binding; 1.  
DR PRINTS; PR00308; ANTIFREEZE1.  
DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; 1.  
DR SMART; SM00395; SANT; 1.

SQ SEQUENCE 3502 AA; 369068 MW; 74C8004F9DA8F8D9 CRC64;  
Query Match 71.1%; Score 32; DB 5; Length 3502;  
Best Local Similarity 66.7%; Pred. No. 5.2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
I :||:||||  
Db 422 GGWVQVSQS 430

RESULT 6  
Q9Y3N1 PRELIMINARY; PRT; 123 AA.  
AC Q9Y3N1;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE DJ695O20A.1 (PEROXISOME PROLIFERATIVE ACTIVATED RECEPTOR, ALPHA)  
DE (FRAGMENT).  
GN PPARA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Phillips S.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.  
DR EMBL; AL049856; CAB42862.2; -.  
DR HSSP; P10826; 1HRA.  
DR InterPro; IPR001628; -.  
DR Pfam; PF00105; zf-C4; 1.  
DR PRINTS; PR0047; STROIDFINGER.  
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;  
KW Zinc-finger.  
FT NON\_TER 123 123  
SQ SEQUENCE 123 AA; 12790 MW; 973587F210F505B8 CRC64;

Query Match 68.9%; Score 31; DB 4; Length 123;  
Best Local Similarity 60.0%; Pred. No. 26;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
I :||:||||  
Db 31 MGNIQEISQS 40

RESULT 7  
Q9VCV0 PRELIMINARY; PRT; 252 AA.  
AC Q9VCV0;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE CG4803 PROTEIN.  
GN CG4803.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fodor C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glisler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirski R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster";  
RT Science 287:2185-2195(2000).  
RL -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC EMBL; AE003740; AAF56055.1; -  
DR FlyBase; FBgn0039015; CG4803.  
DR InterPro; IPR000719; -  
DR InterPro; IPR002290; -  
DR Pfam; PF00069; pkinase; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR SMART; SM00220; S-TKC; 1.  
KW ATP-binding; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 252 AA; 28853 MW; 8E006F2EABB68D17 CRC64;

Query Match 68.9%; Score 31; DB 5; Length 252;  
Best Local Similarity 66.7%; Pred. No. 55;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
DB 155 GTVVDLSQS 163

RESULT 8  
ID O45782 PRELIMINARY; PRT; 394 AA.  
AC O45782;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
DE T12D8.4 PROTEIN.  
GN T12D8.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.

RA McMurray A.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Lalister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smalldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Winkstock L., Wilkinon-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans";  
RL Nature 368:32-38(1994).  
DR EMBL; Z81120; CAB03344.1; -  
SQ SEQUENCE 394 AA; 44031 MW; 35FE9CEBF AAA028C CRC64;

Query Match 68.9%; Score 31; DB 5; Length 394;  
Best Local Similarity 77.8%; Pred. No. 88;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
DB 302 IGSIVQLSQ 310

RESULT 9  
QY14F5 PRELIMINARY; PRT; 436 AA.  
ID QY14F5  
AC QY14F5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE C4-DICARBOXYLATE TRANSPORT PROTEIN.  
DE DCTA OR PA1183.  
GN Pseudomonas aeruginosa  
OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004548; AAG04572.1; -  
DR InterPro; IPR001991; -  
DR Pfam; PF00375; SDF; 1.  
DR PRINTS; PR00173; EDTRNSPORT.  
DR PROSITE; PS00714; NA-DICARBOXYL-SYMP-2; 1.  
SQ SEQUENCE 436 AA; 45993 MW; DCA564DB31A5AC07 CRC64;

Query Match 68.9%; Score 31; DB 2; Length 436;  
Best Local Similarity 66.7%; Pred. No. 98;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
DB 214 VGSIVQLSQ 222



```
RESULT 10
Q9N135
ID Q9N135 PRELIMINARY; PRT; 468 AA.
AC Q9N135;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (FRAGMENT).
GN PPARALPHA.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Sundvold H., Grindflek E., Lien S.;
RT "Peroxisome proliferator activated receptor alpha (PPAR) in swine:
RL detection of an alternative transcript.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
DR EMBL; AF228696; AAF73404.1; -.
DR InterPro; IPR000536; -.
DR InterPro; IPR000741; -.
DR InterPro; IPR001628; -.
DR InterPro; IPR001723; -.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODOM; PD001128; -. 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR SMART; SM00430; HOLI; 1.
DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
FT NON_TER 468 468
SQ SEQUENCE 468 AA; 52011 MW; 5812001D8F44BAAA CRC64;

Query Match 68.9%; Score 31; DB 6; Length 468;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
||:| :|||
Db 31 MGTIQEISQS 40

RESULT 11
Q9N129
ID Q9N129 PRELIMINARY; PRT; 470 AA.
AC Q9N129;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (FRAGMENT).
GN PPARALPHA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Sundvold H., Frengen A.B., Lien S.;
RT "cDNA cloning and expression pattern of the bovine peroxisome
RL proliferator activated receptor (PPAR) alpha and -beta.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
DR EMBL; AF229356; AAF73405.1; -.
DR InterPro; IPR000536; -.

Query Match 68.9%; Score 31; DB 6; Length 468;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
||:| :|||
Db 31 MGTIQEISQS 40

RESULT 12
Q9KWA5
ID Q9KWA5 PRELIMINARY; PRT; 500 AA.
AC Q9KWA5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE RIORF79 PROTEIN.
GN RIORF79.
OS Agrobacterium rhizogenes.
OG Plasmid pRil724.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRil724, by the construction of its physical map and library.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT "Genome structure of Ri plasmid (1):Construction of linking library
RT and physical map of pRil724 in Japanese Agrobacterium.";
RL Nucleic Acids Symp. Ser. 39:189-190(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA Yoshida K.;
RT "The complete nucleotide sequence of a Ri (root inducing) plasmid
RT indicates its chimerical structure between Ti and Sym plasmids.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRil724, by the construction of its physical map and library.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RX MEDLINE=20241294; PubMed=10780382;
RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA Yoshida K.;
RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
```

```
DR InterPro; IPR001628; -.
DR InterPro; IPR001723; -.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR SMART; SM00430; HOLI; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
FT NON_TER 470 470
SQ SEQUENCE 470 AA; 52436 MW; D205B978EC120A4A CRC64;

Query Match 68.9%; Score 31; DB 6; Length 470;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
||:| :|||
Db 31 MGTIQEISQS 40

RESULT 12
Q9KWA5
ID Q9KWA5 PRELIMINARY; PRT; 500 AA.
AC Q9KWA5;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE RIORF79 PROTEIN.
GN RIORF79.
OS Agrobacterium rhizogenes.
OG Plasmid pRil724.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRil724, by the construction of its physical map and library.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT "Genome structure of Ri plasmid (1):Construction of linking library
RT and physical map of pRil724 in Japanese Agrobacterium.";
RL Nucleic Acids Symp. Ser. 39:189-190(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA Yoshida K.;
RT "The complete nucleotide sequence of a Ri (root inducing) plasmid
RT indicates its chimerical structure between Ti and Sym plasmids.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRil724, by the construction of its physical map and library.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724; PLASMID=PR11724;
RX MEDLINE=20241294; PubMed=10780382;
RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA Yoshida K.;
RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
```

RT its flanking regions of pR1724 in Japanese Agrobacterium  
 RT rhizogenes.;  
 RL Nucleic Acids Symp. Ser. 42:67-68(1999).  
 DR EMBL; AB039932; BAA97790.1; -;  
 DR EMBL; AP002086; BAB16198.1; -;  
 DR InterPro; IPR000577; -;  
 DR Pfam; PF00370; FGGY; 1.  
 DR PROSITE; PS00933; FGGY\_KINASES\_1; 1.  
 DR PROSITE; PS00445; FGGY\_KINASES\_2; UNKNOWN\_1.  
 KW Plasmid.  
 SQ SEQUENCE 500 AA; 54183 MW; C2134B8EE931E2F6 CRC64;

Query Match 68.9%; Score 31; DB 2; Length 500;  
 Best Local Similarity 50.0%; Pred. No. 1.le+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 Db 426 LGSVIERQS 435

RESULT 13  
 ID Q9NSZ9 PRELIMINARY; PRT; 617 AA.  
 AC Q9NSZ9;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 67.3 KDA PROTEIN (FRAGMENT).  
 GN DKFZP434G0812.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RA Duesterhoeft A.; Lauber J.; Mewes H.W.; Gassenhuber J.; Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL137619; CAB70845.1; -;  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 617 AA; 67300 MW; F5FB29BB2A59872D CRC64;

Query Match 68.9%; Score 31; DB 4; Length 617;  
 Best Local Similarity 60.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 Db 275 LGSLEQLSQA 284

RESULT 14  
 ID Q9QJC6 PRELIMINARY; PRT; 68 AA.  
 AC Q9QJC6;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE ENVELOPE PROTEIN (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=151N;  
 RA Lopez-Galindez C.;  
 RT "Genetic analysis of HIV-1 Spanish isolates."  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF152836; AAF08474.1; -;  
 DR InterPro; IPR000777; -;  
 DR Pfam; PF00516; GP120; 1.  
 KW Envelope protein.  
 FT NON\_TER 1  
 FT NON\_TER 68  
 SQ SEQUENCE 68 AA; 7538 MW; 7680F5F1734650C7 CRC64;

Query Match 66.7%; Score 30; DB 14; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IVQLSQS 10  
 Db 3 IVQLSQS 9

RESULT 15  
 ID Q9NM69 PRELIMINARY; PRT; 148 AA.  
 AC Q9NM69;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE POSSIBLE ZWH18.1 (FRAGMENT).  
 GN LM26.223.  
 OS Leishmania major.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRIEDLIN;  
 RA Murphy L.; Quail M.; Harris D.; Rajandream M.; Ivens A.; Barrell B.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL160493; CAB97841.1; -;  
 FT NON\_TER 148  
 SQ SEQUENCE 148 AA; 16560 MW; C942B83C686D1162 CRC64;

Query Match 66.7%; Score 30; DB 5; Length 148;  
 Best Local Similarity 66.7%; Pred. No. 53;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
 Db 46 MGSILQLME 54

RESULT 16  
 ID Q9RSZ9 PRELIMINARY; PRT; 161 AA.  
 AC Q9RSZ9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE TRANSCRIPTION ELONGATION FACTOR.  
 GN DE1970.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O.; Eisen J.A.; Heidelberg J.F.; Hickey E.K.; Peterson J.D.;  
 RA Dodson R.S.; Haft D.H.; Gwinn M.L.; Nelson W.C.; Richardson D.L.;  
 RA Moffat K.S.; Qin H.; Jiang L.; Pamphile W.; Crosby M.; Shen M.;  
 RA Vamathevan J.J.; Lam P.; McDonald L.; Utterback T.; Zalewski C.;  
 RA Makarova K.S.; Aravind L.; Daly M.J.; Minton K.W.; Fleischmann R.D.;  
 RA Ketchum K.A.; Nelson K.E.; Salzberg S.; Smith H.O.; Venter J.C.;  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE002035; AAF11523.1; -.  
DR TIGR; DR1970; -.  
DR InterPro; IPR001437; -.  
DR Pfam; PF01272; GreA\_GreB; 1.  
DR ProDom; PD004918; -. 1.  
KW Elongation factor.  
SQ SEQUENCE 161 AA; 18064 MW; 6490C1921015997B CRC64;

Query Match 66.7%; Score 30; DB 2; Length 161;  
Best Local Similarity 75.0%; Pred. No. 58;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLS 8  
Db 88 LGAIVQLS 95

RESULT 17  
O63327 ID O63327 PRELIMINARY; PRT; 170 AA.  
AC O63327;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).  
GN ND2.  
OS Bufo verrucosissimus.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.  
OX NCBI\_TaxID=61429;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPECIMEN NUMBER: MVZ218739;  
RX MEDLINE=98140324; PubMed=9479697;  
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
RA Tuniyev B.S., Papenfuss T.J.;  
RT "Phylogenetic relationships of toads in the Bufo bufo species group  
from the eastern escarpment of the Tibetan Plateau: a case of  
vicariance and dispersal."  
RL Mol. Phylogenet. Evol. 9:80-87(1998).  
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS  
CHAINS.  
DR EMBL; AF004526; AAD05131.1; -.  
DR InterPro; IPR001750; -.  
DR Pfam; PF00361; oxidored\_q1; 1.  
KW Mitochondrion; NAD; Oxidoreductase.  
FT NON\_TER 170 170  
SQ SEQUENCE 170 AA; 18220 MW; 9CFD07402DBE6503 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;  
Best Local Similarity 60.0%; Pred. No. 62;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLS 10  
Db 139 MALLIQLS 148

RESULT 18  
O63329 ID O63329 PRELIMINARY; PRT; 170 AA.  
AC O63329;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).  
GN ND2.  
OS Bufo andrewsi.

OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.  
OX NCBI\_TaxID=61428;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPECIMEN NUMBER: MVZ216648;  
RX MEDLINE=98140324; PubMed=9479697;  
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
RA Tuniyev B.S., Papenfuss T.J.;  
RT "Phylogenetic relationships of toads in the Bufo bufo species group  
from the eastern escarpment of the Tibetan Plateau: a case of  
vicariance and dispersal."  
RL Mol. Phylogenet. Evol. 9:80-87(1998).  
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS  
CHAINS.  
DR EMBL; AF004529; AAD05137.1; -.  
DR InterPro; IPR001750; -.  
DR Pfam; PF00361; oxidored\_q1; 1.  
KW Mitochondrion; NAD; Oxidoreductase.  
FT NON\_TER 170 170  
SQ SEQUENCE 170 AA; 18225 MW; 2E3705086E28F668 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;  
Best Local Similarity 60.0%; Pred. No. 62;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLS 10  
Db 139 MALLIQLS 148

RESULT 19  
O63331 ID O63331 PRELIMINARY; PRT; 170 AA.  
AC O63331;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).  
GN ND2.  
OS Bufo gargarizans.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.  
OX NCBI\_TaxID=30331;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPECIMEN NUMBER: CAS194318;  
RX MEDLINE=98140324; PubMed=9479697;  
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,  
RA Tuniyev B.S., Papenfuss T.J.;  
RT "Phylogenetic relationships of toads in the Bufo bufo species group  
from the eastern escarpment of the Tibetan Plateau: a case of  
vicariance and dispersal."  
RL Mol. Phylogenet. Evol. 9:80-87(1998).  
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS  
CHAINS.  
DR EMBL; AF004530; AAD05139.1; -.  
DR InterPro; IPR001750; -.  
DR Pfam; PF00361; oxidored\_q1; 1.  
KW Mitochondrion; NAD; Oxidoreductase.  
FT NON\_TER 170 170  
SQ SEQUENCE 170 AA; 18340 MW; 22F619F1E35B5458 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;  
Best Local Similarity 60.0%; Pred. No. 62;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLS 10  
Db 139 MALLIQLS 148

```
Db 139 MALLIQLSQS 148
RESULT 20
ID O63332 PRELIMINARY; PRT; 170 AA.
AC O63332;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
OX NCBI_TaxID=30331;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPECIMEN NUMBER: CAS178100;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tuniyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004531; AAD05141.1; -
DR InterPro; IPR001750; -
DR Pfam; PF00361; Oxidored_ql; 1.
DR Mitochondrion; NAD; Oxidoreductase.
KW NON_TER 170
SQ SEQUENCE 170 AA; 18266 MW; 384DB9D87BEA7DCB CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. NO. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 139 MALLIQLSQS 148

RESULT 21
ID O63333 PRELIMINARY; PRT; 170 AA.
AC O63333;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
OX NCBI_TaxID=30331;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPECIMEN NUMBER: MVZ211365;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tuniyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004533; AAD05145.1; -
```

```
DR InterPro; IPR001750; -
DR Pfam; PF00361; Oxidored_ql; 1.
KW Mitochondrion; NAD; Oxidoreductase.
FT NON_TER 170
SQ SEQUENCE 170 AA; 18284 MW; 38417E1C0589C460 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. NO. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 139 MALLIQLSQS 148

RESULT 22
ID O63957 PRELIMINARY; PRT; 170 AA.
AC O63957;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
OX NCBI_TaxID=30331;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tuniyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004535; AAD05149.1; -
DR EMBL; AF004532; AAD05143.1; -
DR EMBL; AF004534; AAD05147.1; -
DR InterPro; IPR001750; -
DR Pfam; PF00361; Oxidored_ql; 1.
KW Mitochondrion; NAD; Oxidoreductase.
FT NON_TER 170
SQ SEQUENCE 170 AA; 18298 MW; 385712C2CA416760 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. NO. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 139 MALLIQLSQS 148

RESULT 23
ID O64012 PRELIMINARY; PRT; 170 AA.
AC O64012;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo andrewsi.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



```
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
OX NCBI_TaxID=61428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPECIMEN NUMBER: CAS194477, SPECIMEN NUMBER: CAS194888;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tuniyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal.";
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004528; AAD05135.1; -.
DR EMBL; AF004527; AAD05133.1; -.
DR InterPro; IPR001750; -.
DR Pfam; PF00361; oxidored_ql; 1.
KW Mitochondrion; NAD; Oxidoreductase.
FT NON_TER 170 170
SQ SEQUENCE 170 AA; 18316 MW; 352DBF02DB58CCD2 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 139 MALLIQLSQS 148

RESULT 24
O21236 PRELIMINARY; PRT; 193 AA.
AC O21236;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE RIBOSOMAL PROTEIN L10.
GN RPL10.
OS Reclinomonas americana.
OG Mitochondrion.
OC Eukaryota; core jakobids; Reclinomonas.
OX NCBI_TaxID=48483;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50394;
RX MEDLINE=97311393; PubMed=9168110;
RA Lang B.F., Burger G., O'Kelly C.J., Cedergren R., Golding G.B.,
RA Lemieux C., Sankoff D., Turmel M., Gray M.W.;
RT "An ancestral mitochondrial DNA resembling a eubacterial genome in
RT miniature.";
RL Nature 387:493-497(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50394;
RA Lang B.F., Burger G.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007261; AAD11863.1; -.
DR InterPro; IPR001790; -.
DR Pfam; PF00466; Ribosomal_L10; 1.
KW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 193 AA; 22448 MW; B1DFCE7D3D7C023F CRC64;

Query Match 66.7%; Score 30; DB 8; Length 193;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SIVQLSQS 10
Db 161 SVVQLTQS 168
```

```
RESULT 25
Q9HWT0
ID Q9HWT0 PRELIMINARY; PRT; 241 AA.
AC Q9HWT0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PROBABLE SHORT-CHAIN DEHYDROGENASE.
GN PA4098.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004826; AAG07485.1; -.
DR InterPro; IPR002198; -.
DR InterPro; IPR002347; -.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00678; adh_short_C2; 1.
DR PRINTS; PR00081; GDRDH.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
SQ SEQUENCE 241 AA; 25622 MW; 879AAD48E36F19CE CRC64;

Query Match 66.7%; Score 30; DB 2; Length 241;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
Db 151 GAIVQLTRS 159

RESULT 26
Q9KJP3
ID Q9KJP3 PRELIMINARY; PRT; 261 AA.
AC Q9KJP3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE 4-OXALOCROTONATE DECARBOXYLASE-LIKE PROTEIN.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DZ2;
RX MEDLINE=20032045; PubMed=10564511;
RA Cho K., Zusman D.R.;
RT "Sporulation timing in Myxococcus xanthus is controlled by the espAB
RT locus.";
RL Mol. Microbiol. 34:714-725(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DZ2;
RX MEDLINE=20032045; PubMed=10564511;
RA Cho K., Zusman D.R.;
RT "Sporulation timing in Myxococcus xanthus is controlled by the espAB
RT locus.";
RL Mol. Microbiol. 34:714-725(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DZ2;
RA Cho K., Zusman D.R.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
```

DR EMBL; AF163841; AAF87924.1; --  
DR InterPro; IPR002607; --  
DR Pfam; PF01689; Hydratase-decarb; 1.  
SQ SEQUENCE 261 AA; 28117 MW; 8562E13FC522FA9A CRC64;

Query Match 66.7%; Score 30; DB 2; Length 261;  
Best Local Similarity 62.5%; Pred. No. 97;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 9  
DB 92 GGVQLSQ 99

RESULT 27  
Q9IPS9 PRELIMINARY; PRT; 269 AA.  
AC Q9IPS9  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE COAT PROTEIN (FRAGMENT).  
GN CPl.  
OS Tobacco necrosis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OC Necrovirus.  
OX NCBI\_TaxID-12054;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TOYAMA;  
RA Sasaki K., Takahashi Y., Oh-oka H., Umeoka T., Oda Y., Fukuyama K.;  
RT "Amino Acid Sequence of the Coat Protein of a Tobacco Necrosis Virus  
Strain Isolated as a Cause of Tulip Necrotic Disease";  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB037115; BAB01493.1; --  
DR InterPro; IPR000937; --  
DR Pfam; PF00729; Viral\_coat; 1.  
DR PRINTS; PR00233; ICOSAHEDRAL.  
FT NON\_TER 1  
FT TER 269  
SQ SEQUENCE 269 AA; 29145 MW; FDA4EF8E26D69C9C CRC64;

Query Match 66.7%; Score 30; DB 14; Length 269;  
Best Local Similarity 77.8%; Pred. No. 1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 10  
DB 160 GSRVQLSQ 168

RESULT 28  
Q88611 PRELIMINARY; PRT; 277 AA.  
AC Q88611  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
DE COAT PROTEIN.  
OS Tobacco necrosis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
OC Necrovirus.  
OX NCBI\_TaxID-12054;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhang L.;  
RL Thesis (1993), University of Nebraska, Lincoln, USA.  
DR EMBL; L04261; CAB23719.1; --  
DR HSSP; P03607; 4SBV.  
DR InterPro; IPR000937; --  
DR Pfam; PF00729; Viral\_coat; 1.

DR PRINTS; PR00233; ICOSAHEDRAL.  
DR Coat protein.  
SQ SEQUENCE 277 AA; 30275 MW; 99B81D725B614609 CRC64;

Query Match 66.7%; Score 30; DB 14; Length 277;  
Best Local Similarity 77.8%; Pred. No. 1e+02;  
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 2 GSIVQLSQS 10  
DB 168 GSRVQLSQ 176

RESULT 29  
Q9PJ60 PRELIMINARY; PRT; 280 AA.  
AC Q9PJ60  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PUTATIVE FLAGELLAR MOTOR SWITCH PROTEIN.  
GN FLIY OR CJ0059.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID-197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RX MEDLINE-20150912; PubMed-10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
DR EMBL; AL139074; CAB72546.1; --  
DR InterPro; IPR001172; --  
DR InterPro; IPR001543; --  
DR Pfam; PF01052; SPOA; 1  
DR PRINTS; PR00956; FLGMOTORFLIN.  
SQ SEQUENCE 280 AA; 31023 MW; 060AC48D485B11B7 CRC64;

Query Match 66.7%; Score 30; DB 2; Length 280;  
Best Local Similarity 55.6%; Pred. No. 1e+02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
DB 227 IGSVVQLNQ 235

RESULT 30  
O48641 PRELIMINARY; PRT; 327 AA.  
AC O48641  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PROTEIN PHOSPHATASE 1 CATALYTIC SUBUNIT.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID-3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Mitsuoka N., Okumura S., Shibata D.;

RL Soil Sci. Plant Nutrition 43:971-974(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=20181125; PubMed=10718197;

RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,

RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence

RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC

RT clones.";

RL DNA Res. 7:31-63(2000).

CC -!- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +

CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC) (BY

CC SIMILARITY).

CC -!- SIMILARITY: TO SERINE/THREONINE SPECIFIC PROTEIN PHOSPHATASE (OR

CC PPP) FAMILY.

DR EMBL; AB000094; BAA24283.1; -.

DR EMBL; AB025638; BAA97417.1; -.

DR HSSP; P08129; 1FJM.

DR Mendel; 24637; Arath;1262;24637.

DR InterPro; IPR000934; -.

DR Pfam; PF00149; STphosphatase; 1.

DR PRINTS; PR00114; STPHPTASE.

DR PROSITE; PS00125; SER\_THR\_PHOSPHATASE; UNKNOWN\_1.

DR SMART; SM00156; PP2AC; 1.

KW Hydrolase; Iron; Manganese.

SQ SEQUENCE 327 AA; 37281 MW; CBEEFA363F8C635A CRC64;

Query Match 66.7%; Score 30; DB 10; Length 327;

Best Local Similarity 66.7%; Pred. No. 1.2e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

DB 21 GKIVQLSET 29

RESULT 31

Q9P367 PRELIMINARY; PRT; 364 AA.

AC Q9P367;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE GRR1 RELATED PROTEIN (FRAGMENT).

GN 8D4.260.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (APR-2000) to the EMBL/GenBank/DBBJ databases.

DR EMBL; AL353819; CAB88568.1; -.

DR InterPro; IPR003592; -.

DR SMART; SM00370; LRR; 1.

FT NON\_TER 364 364

SQ SEQUENCE 364 AA; 40305 MW; 069A34B167A461CC CRC64;

Query Match 66.7%; Score 30; DB 3; Length 364;

Best Local Similarity 60.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10

: |:|:|:|

Db 252 IASLVQLAQS 261

RESULT 32

Q9SV55 PRELIMINARY; PRT; 425 AA.

AC Q9SV55;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE HYPOTHETICAL 44.9 KDA PROTEIN.

GN F25O24.30 OR AT4G28910.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,

RA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;

RL Submitted (JUN-1999) to the EMBL/GenBank/DBBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (AUG-1999) to the EMBL/GenBank/DBBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,

RA Mewes H.W., Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBBJ databases.

DR EMBL; AL078469; CAB43905.1; -.

DR EMBL; AL161573; CAB81479.1; -.

KW Hypothetical protein.

SQ SEQUENCE 425 AA; 44938 MW; 1796FB2FAAB0FB89 CRC64;

Query Match 66.7%; Score 30; DB 10; Length 425;

Best Local Similarity 77.8%; Pred. No. 1.6e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

: |:|:|:|

Db 279 GGIVALSQS 287

RESULT 33

Q9RK75 PRELIMINARY; PRT; 473 AA.

AC Q9RK75;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE PUTATIVE HYDROLASE.

GN SCF11.15.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBBJ databases.

```
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=A3(2); PubMed=8843436;
RC MEDLINE=9700351; Kleser H.M., Denapaita D., Eichner A., Cullum J.,
RA Redenbach M., Hopwood D.A.;
RT Kinashi H., and a detailed genetic and physical map for
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL132662; CAB59592.1; -
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; -
DR Pfam; PF00942; CBD_3; 1.
KW Hydrolase.
SQ SEQUENCE 473 AA; 50526 MW; 47AE1E4CDFCC264D CRC64;

Query Match 66.7%; Score 30; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SIVQLSQ 9
DB 67 SIVQLSQ 73

RESULT 34
Q9PDU7 PRELIMINARY; PRT; 503 AA.
ID Q9PDU7
AC Q9PDU7
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CARBOXYPEPTIDASE RELATED PROTEIN.
GN XFL282.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9A5C;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barrios M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE003962; AAF84091.1; -
DR InterPro; IPR000379; -
```

```
SQ SEQUENCE 503 AA; 55835 MW; 049EC5CA922F44EB CRC64;

Query Match 66.7%; Score 30; DB 2; Length 503;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
DB 417 GSIVQLSQS 425

RESULT 35
Q9YD39 PRELIMINARY; PRT; 539 AA.
ID Q9YD39
AC Q9YD39
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 61.4 KDA PROTEIN APE1073.
GN APE1073.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K1; PubMed=10382966;
RC MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oquchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY.
DR EMBL; AP000060; BAA80058.1; -
DR InterPro; IPR001161; -
DR InterPro; IPR001410; -
DR InterPro; IPR001650; -
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR PRINTS; PR00851; XRODRMPGMNTB.
DR SMART; SM00490; HELIC; 1.
KW ATP-binding; Helicase; Hypothetical protein.
SQ SEQUENCE 539 AA; 61395 MW; 3E8C78F07C0DD0F4 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 539;
Best Local Similarity 70.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
DB 396 MKEIVQLSES 405

RESULT 36
Q9VGP3 PRELIMINARY; PRT; 587 AA.
ID Q9VGP3
AC Q9VGP3
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CG6723 PROTEIN.
GN CG6723.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
```



OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Stroh R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissensbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003691; AAF54633.1; -.  
DR FlyBase; FBgn0037895; CG6723.  
DR InterPro; IPR001734; -.  
DR Pfam; PF00474; SSF; 1.  
DR PROSITE; PS0283; NA\_SOLUT\_SYMP\_3; 1.  
SQ SEQUENCE 587 AA; 64403 MW; 946FEBEB30C97E1E CRC64;

Query Match 66.7%; Score 30; DB 5; Length 587;  
Best Local Similarity 50.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10  
Db 410 MGVVLQLTMS 419  
II:::II: I  
RESULT 37  
Q94486  
ID Q94486 PRELIMINARY; PRT; 624 AA.  
AC Q94486;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last annotation update)  
DE ORF DG1040 (FRAGMENT).  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=AX4;  
RA Loomis W.F.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; U66913; AAB07544.1; -.  
FT NON\_TER 624 624  
SQ SEQUENCE 624 AA; 70445 MW; 50343A91ECB0CB37 CRC64;

Query Match 66.7%; Score 30; DB 5; Length 624;  
Best Local Similarity 50.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10  
Db 263 LNTLIQLSQS 272  
:::IIII I

RESULT 38  
Q9UFJ0  
ID Q9UFJ0 PRELIMINARY; PRT; 662 AA.  
AC Q9UFJ0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE HYPOTHETICAL 70.3 KDA PROTEIN (FRAGMENT).  
GN DKF2P434O225.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AL117647; CAB56029.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 662 AA; 70338 MW; D2B3562B55ACFF18 CRC64;

Query Match 66.7%; Score 30; DB 4; Length 662;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSIVQLS 8  
Db 590 GSVVQLS 596  
II:IIII

RESULT 39  
Q9NQF8  
ID Q9NQF8 PRELIMINARY; PRT; 703 AA.  
AC Q9NQF8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE DJ1184F4.1 (KIAA0978) (FRAGMENT).  
GN DJ1184F4.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Barlow K.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AL034550; CAC00607.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 703 AA; 74741 MW; 8EE4B3F35672F012 CRC64;

Query Match 66.7%; Score 30; DB 4; Length 703;

```

Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSVVQLS 8
Db 631 GSVVQLS 637

RESULT 40
P74111 PRELIMINARY; PRT; 750 AA.
AC P74111;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SENSORY TRANSDUCTION HISTIDINE KINASE.
GN SLR1969.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8905231;
RX MEDLINE=97061201; Tanaka A., Asamizu E., Nakamura Y.,
RA Kaneko T., Sato S., Kotani H., Sugiyama M., Sasamoto S., Kimura T.,
RA Miyajima N., Hirose M., Saito M., Nakazaki N., Naruo K., Okumura S.,
RA Hosouchi T., Matsuno A., Muraki A., Watanabe A., Yamada M., Yasuda M.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RT DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS. HISTIDINE
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC EMBL; D90912; BAA18197.1; -
DR InterPro; IPR000410; -
DR InterPro; IPR001294; -
DR InterPro; IPR001789; -
DR InterPro; IPR003018; -
DR Pfam; PF00072; response_reg; 1.
DR Pfam; PF00512; signal; 1.
DR Pfam; PF01590; GAF; 1.
DR PROSITE; PS50046; PHYTOCHROME_2; 1.
DR SMART; SM00448; REC; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase.
SQ SEQUENCE 750 AA; 83989 MW; D3DCB338F421896A CRC64;

Query Match 66.7%; Score 30; DB 2; Length 750;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
Db 437 INSIIQLSQ 445

RESULT 41
Q27537 PRELIMINARY; PRT; 816 AA.
AC Q27537;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SIMILARITY TO DNA TOPOISOMERASE II.
GN ZK1127.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton A., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierri-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohldman P., III of C.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS (BY SIMILARITY).
CC -1- SIMILARITY: TO DNA TOPOISOMERASE II FAMILY.
DR EMBL; U58758; AAB93429.1; -
DR HSP; P06786; IBCW
DR InterPro; IPR001241; -
DR InterPro; IPR002205; -
DR Pfam; PF00204; DNA_topoisoiI; 1.
DR Pfam; PF00521; DNA_topoisoiV; 1.
DR ProDom; PD000616; -; 1.
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
DR SMART; SM00434; TOP4C; 1.
KW ATP-binding; DNA-binding; Isomerase; Topoisomerase.
SQ SEQUENCE 816 AA; 93029 MW; B3A0A6C61D061AEB CRC64;

```

```

Query Match 66.7%; Score 30; DB 5; Length 816;
Best Local Similarity 66.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
Db 373 MGTIVNLAQ 381

RESULT 42
Q9VKQ6 PRELIMINARY; PRT; 944 AA.
ID Q9VKQ6
AC Q9VKQ6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)

```

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE CG6729 PROTEIN.  
GN CG6729.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003629; AAF53005.1; -.  
DR FLYBase; FBgn0032296; CG6729.  
SQ SEQUENCE 944 AA; 106827 MW; E02D6405B35645BF CRC64;

Query Match 66.7%; Score 30; DB 5; Length 944;  
Best Local Similarity 75.0%; Pred. No. 3.7e+02;  
Matches 9; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 1 MGSIV--QLSQS 10  
ID Q9LJX4 PRELIMINARY; PRT; 962 AA.  
AC Q9LJX4;  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE RNA BINDING PROTEIN-LIKE.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX PubMed=10907853;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,  
RT TAC and BAC clones."  
RL DNA Res. 7:217-221(2000).  
DR EMBL; AP000383; BAB01884.1; -.  
DR InterPro; IPR001313; -.  
DR Pfam; PF00806; PUF; 8.  
SQ SEQUENCE 962 AA; 107201 MW; E95B8CBEB375D2F2 CRC64;

Query Match 66.7%; Score 30; DB 10; Length 962;  
Best Local Similarity 62.5%; Pred. No. 3.8e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIIVQLSQ 9  
ID 849 GNVVQMSQ 856

RESULT 44  
Q9ZW02 PRELIMINARY; PRT; 964 AA.  
ID Q9ZW02  
AC Q9ZW02;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
DE PUMILIO-LIKE PROTEIN.  
GN F16P2.48.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;  
RT "Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence."  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004561; AAC95220.1; -.  
DR InterPro; IPR001313; -.  
DR Pfam; PF00806; PUF; 8.  
DR SMART; SM00025; Pumilio; 1.  
SQ SEQUENCE 964 AA; 105986 MW; 40E3FFB1B5A6E43A CRC64;

Query Match 66.7%; Score 30; DB 10; Length 964;  
Best Local Similarity 75.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIIVQLSQ 9  
ID 846 GKIVQMSQ 853

RESULT 45  
Q9M6S0

```
ID Q9M6S0 PRELIMINARY; PRT; 966 AA.
AC Q9M6S0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUMILIO DOMAIN-CONTAINING PROTEIN PPD1.
GN PPD1.
OS Populus tremula x Populus tremuloides.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=47664;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhalerao R.P.; Amini B.; Lundeborg J.; Sandberg G.;
RT "A pumilio RNA binding domain containing protein is expressed in
RT vascular tissue of Hybrid Aspen.";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF153276; AAF71823.1; -
DR InterPro; IPR001313; -
DR InterPro; IPR001412; -
DR Pfam; PF00806; PUF; 8.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
SQ SEQUENCE 966 AA; 105324 MW; 728CF2320C28E652 CRC64;
```

Query Match 66.7%; Score 30; DB 10; Length 966;  
Best Local Similarity 75.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 9  
Db 884 GKIVQMSQ 891  
| | | | |

Search completed: June 28, 2001, 11:55:40  
Job time: 161 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:54:15 ; Search time 66.55 Seconds  
(without alignments)  
24.596 Million cell updates/sec

Title: US-09-439-313-566  
Perfect score: 146  
Sequence: 1 VGEGLYQGVPAEPGTEARRHYDEGVR 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	100.0	553	AAW71869	Amino acid encoded
2	146	100.0	553	AAW69385	Prostate tumour sp
3	146	100.0	553	AAW28527	Protein encoded by
4	146	100.0	553	AAW82002	Human immunogenic
5	131	89.7	27	AAW85071	PS108 protein deri
6	131	89.7	27	AAW29271	Human prostate-rel
7	131	89.7	255	AAW85068	Protein encoded by
8	131	89.7	255	AAW29268	Human prostate-rel
9	63	43.2	742	AAW94977	Human secreted pro
10	50	34.2	27	AAW09155	Hepatitis GB virus
11	50	34.2	3164	AAW94346	Hepatitis GB virus

12	48	32.9	113	21	AAG32765	Zea mays protein f
13	48	32.9	123	21	AAG27261	Zea mays protein f
14	48	32.9	563	20	AAW41125	Wheat sucrose tran
15	47	32.2	90	21	AAW53798	Amino acids 145-23
16	47	32.2	439	21	AAW53797	Amino acid sequenc
17	47	32.2	913	21	AAG14138	Arabidopsis thalia
18	47	32.2	916	21	AAG37416	Arabidopsis thalia
19	47	32.2	994	21	AAG14137	Arabidopsis thalia
20	47	32.2	997	21	AAG37415	Arabidopsis thalia
21	47	32.2	1098	21	AAG14136	Arabidopsis thalia
22	47	32.2	1101	21	AAG37414	Arabidopsis thalia
23	46.5	31.8	140	19	AAW42074	The amino acid seq
24	46.5	31.8	140	19	AAW41600	Secreted protein C
25	46.5	31.8	240	19	AAW41601	Secreted protein C
26	46.5	31.8	917	20	AAW03096	A human FAN (facto
27	46	31.5	139	19	AAW77649	Phosphoribosylform
28	46	31.5	499	18	AAW36183	Monkey p53 tumour
29	46	31.5	529	16	AAW72328	Laccase RSlac3. R
30	46	31.5	529	18	AAW16301	Rhizoctonia solani
31	46	31.5	529	19	AAW76310	Rhizoctonia solani
32	46	31.5	529	19	AAW60878	Rhizoctonia solani
33	46	31.5	529	21	AAW50728	R. solani laccase
34	46	31.5	541	21	AAW25538	Eucalyptus grandis
35	46	31.5	566	20	AAW86348	Human multiple end
36	46	31.5	610	19	AAW29749	Tumor suppressor p
37	46	31.5	619	21	AAW40554	Human ORFX ORF318
38	46	31.5	637	18	AAW36182	Monkey p53 tumour
39	46	31.5	1088	17	AAW88634	Plasmid pASK75 ope
40	46	31.5	1277	15	AAW52702	Plasmid pASK60-Str
41	45	30.8	170	21	AAW36228	Rat Fas receptor
42	45	30.8	411	21	AAW87981	Pseudomonas sp. HR
43	45	30.8	586	21	AAW87980	Pseudomonas sp. HR
44	45	30.8	673	21	AAW87979	Pseudomonas sp. HR
45	44.5	30.5	157	22	AAW19929	Human oxidoreducta
46	44	30.1	254	21	AAW06572	Arabidopsis thalia
47	44	30.1	270	21	AAW06571	Arabidopsis thalia
48	44	30.1	340	21	AAW43972	Human cancer assoc
49	44	30.1	363	21	AAW96748	A. terreus LovC de
50	44	30.1	406	21	AAW06570	Arabidopsis thalia
51	43.5	29.8	344	21	AAW31132	Arabidopsis thalia
52	43.5	29.8	369	21	AAW31131	Arabidopsis thalia
53	43.5	29.8	369	21	AAW39949	Arabidopsis thalia
54	43.5	29.8	394	21	AAW39948	Arabidopsis thalia
55	43.5	29.8	450	21	AAW31130	Arabidopsis thalia
56	43.5	29.8	475	21	AAW39947	Arabidopsis thalia
57	43.5	29.8	4544	15	AAW47861	Alpha 2-Macroglobu
58	43.5	29.8	4544	15	AAW60517	Human alpha-2-MR.
59	43	29.5	113	21	AAW52344	Mutant full-length
60	43	29.5	376	20	AAW50035	Human complement r
61	43	29.5	381	8	AAW70048	Human decay accele
62	43	29.5	381	10	AAW94773	Decay accelerating
63	43	29.5	381	16	AAW66683	Decay accelerating
64	43	29.5	381	18	AAW26317	Human decay accele
65	43	29.5	381	20	AAW31740	Human CD55 and 791

ALIGNMENTS

RESULT	1
AAW71869	
ID	AAW71869 standard; Protein; 553 AA.
XX	
AC	AAW71869;
XX	
DT	06-JAN-1999 (first entry)
XX	
DE	Amino acid encoded by prostate tumour clone L1-12.
XX	
KW	Prostate; cancer; tumour; vaccine; immunogen; clone.
XX	
OS	Homo sapiens.
XX	

PN WO9837093-A2.  
XX  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03492.  
XX  
PR 09-FEB-1998; 98US-0020956.  
XX  
PR 25-FEB-1997; 97US-0806099.  
XX  
PR 01-AUG-1997; 97US-0904804.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI DILLON DC, Xu J;  
XX  
DR WPI; 1998-609886/51.  
XX  
DR N-PSDB; AAV61201.  
XX  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
XX used in a vaccine for the treatment of prostate cancer  
PS Example 1; Page 82-84; 130pp; English.  
XX  
CC The present sequence is an immunogenic portion of a prostate tumour  
CC protein. The immunogen, or the DNA encoding it, can be used as a  
CC vaccine for the treatment of prostate cancer. The immunogen was  
CC isolated from a prostate tumour cDNA library obtained by subtracting  
CC a prostate tumour cDNA expression library with a normal tissue cDNA  
CC library.  
XX  
SQ Sequence 553 AA;  
  
Query Match 100.0%; Score 146; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 4.8e-13;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VEGELYQGVPRAEPGTEARRHYDEGVR 27  
Db ||||||||||||||||||||||||||||  
296 vgeglyqgvpraepgtearrhydegvr 322  
  
RESULT 2  
AAW69385  
ID AAW69385 standard; Protein: 553 AA.  
XX  
AC AAW69385;  
XX  
DT 08-DEC-1998 (first entry)  
XX  
DE Prostate tumour specific gene clone L1-12 protein.  
XX  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
XX therapy.  
XX Homo sapiens.  
OS  
XX WO9837418-A2.  
PN  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03690.  
XX  
PR 09-FEB-1998; 98US-0904809.  
XX  
PR 25-FEB-1997; 97US-0806596.  
XX  
PR 01-AUG-1997; 97US-0904809.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI DILLON DC, Xu J;  
XX  
DR WPI; 1998-480805/41.  
XX  
DR N-PSDB; AAV58586.  
XX

PT Novel human prostate specific tumour protein and fragments - useful  
XX for detecting and treating prostate cancers  
PS Example 1; Page 87-89; 141pp; English.  
XX  
CC This sequence is encoded by a human prostate tumour specific gene, and  
CC can be used in the method of the invention. The method is for detecting  
CC prostate cancer comprising contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC this protein sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
XX  
SQ Sequence 553 AA;  
  
Query Match 100.0%; Score 146; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 4.8e-13;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VEGELYQGVPRAEPGTEARRHYDEGVR 27  
Db ||||||||||||||||||||||||||||  
296 vgeglyqgvpraepgtearrhydegvr 322  
  
RESULT 3  
AAB28527  
ID AAB28527 standard; Protein: 553 AA.  
XX  
AC AAB28527;  
XX  
DT 07-FEB-2001 (first entry)  
XX  
DE Protein encoded by human breast tumour cDNA clone P501S.  
XX  
KW Human; breast tumour antigen; cytostatic; immunotherapy;  
XX breast cancer; vaccine.  
XX Homo sapiens.  
OS  
XX WO200061756-A2.  
PN  
PD 19-OCT-2000.  
XX  
PF 10-APR-2000; 2000WO-US09688.  
XX  
PR 09-APR-1999; 99US-0288950.  
XX  
PR 02-JUL-1999; 99US-0346327.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Reed SG, Xu J, Dillon DC;  
XX  
DR WPI; 2000-638568/61.  
XX  
DR N-PSDB; AAC79473.  
XX  
PT A novel isolated polypeptide comprising an immunogenic portion of a  
XX breast cancer protein useful in the detection and treatment of breast  
XX cancer -  
PS Claim 2; Page 92-93; 95pp; English.  
XX  
CC The present sequence is encoded by a cDNA sequence which was isolated  
CC from a breast tumour cDNA library. It is provided in a specification  
CC relating to compounds for immunotherapy and diagnosis of breast cancer.  
CC Breast tumour antigens and the polynucleotides that encode them may be  
CC used in the production of a pharmaceutical composition to be used in the  
CC treatment of breast cancer. Proliferated T cells and incubated antigen  
CC presenting cells are also required. The polypeptides and polynucleotides  
CC may also be used to produce a vaccine.  
XX

SQ Sequence 553 AA;

Query Match 100.0%; Score 146; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 4.8e-13;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEGGLYQGVPRAPGTEARRHYDEGVR 27  
|||||  
Db 296 vgeglyqgvpraepgtearrhydegvr 322

RESULT 4  
AA82002  
ID AAY82002 standard; Protein; 553 AA.  
XX  
AC AAY82002;  
XX  
DT 13-JUN-2000 (first entry)  
XX  
DE Human immunogenic prostate tumour protein sequence SEQ ID NO:113.  
XX  
DE Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
KW immunogenic; cytostatic; vaccine.  
KW  
XX  
OS Homo sapiens.  
XX  
PN WO200004149-A2.  
XX  
PD 27-JAN-2000.  
XX  
PF 14-JUL-1999; 99WO-US15838.  
XX  
PR 14-JUL-1998; 98US-0115453.  
PR 14-JUL-1998; 98US-0116134.  
PR 23-SEP-1998; 98US-0159812.  
PR 23-SEP-1998; 98US-0159822.  
PR 15-JAN-1999; 99US-0232149.  
PR 15-JAN-1999; 99US-0232880.  
PR 09-APR-1999; 99US-0288946.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;  
XX  
DR WPI; 2000-171268/15.  
XX  
PT New polypeptide useful for treating and diagnosing prostate cancer  
PT comprises an immunogenic portion of prostate tumor protein -  
XX  
PS Claim 3; Page 138-139; 263pp; English.  
XX  
CC The present invention describes isolated polypeptides, comprising an  
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides  
CC and polynucleotides encoding them have cytostatic activity and can be  
CC used in vaccines and in gene therapy. The polypeptides and  
CC polynucleotides encoding them, antigen presenting cells which express  
CC the polypeptides, antibodies against the polypeptides and vaccines  
CC comprising them can be used for inhibiting the development of prostate  
CC cancer in a patient. The polypeptides can be used to generate antibodies  
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of  
CC the polynucleotides encoding the polypeptides can be used as a probe or  
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and  
CC AAY82000 to AAY82020 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 146; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 4.8e-13;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEGGLYQGVPRAPGTEARRHYDEGVR 27  
|||||  
Db 296 vgeglyqgvpraepgtearrhydegvr 322

RESULT 5  
AA85071  
ID AAW85071 standard; Peptide; 27 AA.  
XX  
AC AAW85071;  
XX  
DT 12-FEB-1999 (first entry)  
XX  
DE PS108 protein derived peptide.  
XX  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO9850567-A1.  
XX  
PD 12-NOV-1998.  
XX  
PF 01-MAY-1998; 98WO-US08930.  
XX  
PR 02-MAY-1997; 97US-0850713.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
DR WPI; 1999-034731/03.  
XX  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
XX  
PS Claim 17; Page 101; 122pp; English.  
XX  
CC The present sequence is derived from the PS108 protein. The  
CC specification describes a method for detecting the presence  
CC of a target PS108 polynucleotide in a test sample. The method  
CC comprises contacting the test sample with at least 1 PS108-specific  
CC polynucleotide or complement, and detecting the presence of the target  
CC PS108 polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
XX  
SQ Sequence 27 AA;

Query Match 89.7%; Score 131; DB 20; Length 27;  
Best Local Similarity 100.0%; Pred. No. 2.6e-12;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GLYQGVPRAPGTEARRHYDEGVR 27  
|||||  
Db 1 glyqgvpraepgtearrhydegvr 24

RESULT 6  
AAB29271  
ID AAB29271 standard; Protein; 27 AA.  
XX  
AC AAB29271;  
XX

DT 07-FEB-2001 (first entry)  
XX Human prostate-related PS108 partial protein sequence SEQ ID NO: 39.  
DE Human; prostate cancer; PS108; antibody; tumour; metastasis.  
KW Homo sapiens.  
XX US6130043-A.  
XX 10-OCT-2000.  
XX 01-MAY-1998; 98US-0071710.  
XX 02-MAY-1997; 97US-0850713.  
XX (ABBO ) ABBOTT LAB.  
XX Cohen M, Colpitts TL, Friedman PN, Gordon J, Granados EN;  
PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;  
PI Kratochvil JD, Russell JC, Hodges SC;  
XX WPI; 2000-655655/63.  
XX Methods for detecting target prostate-specific polynucleotides or  
PT diseases of the prostate (e.g. prostate cancer), comprising detecting  
PT the presence of any of PS108 nucleic acid sequences in a test sample -  
XX Example 10; Column 87-88; 55pp; English.  
PS Sequence 27 AA;  
XX  
XX The present invention is related to a number of partial coding and  
CC protein sequences for the human prostate tissue protein PS108. These  
CC sequences can be used in the diagnosis and prognosis of prostate  
CC diseases, particularly prostate cancer. They can also be used to produce  
CC antibodies which can be used in treatment. The present sequence is one  
CC of the PS108 partial protein sequences.  
XX Sequence 27 AA;  
SQ

Query Match 89.7%; Score 131; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 2.6e-12;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GLYQGVPRAPGTEARRHYDEGVR 27  
Db 1 glyqgvpraepgtearrhydegvr\*24

RESULT 7  
AAW85068  
ID AAW85068 standard; Protein; 255 AA.  
XX  
AC AAW85068;  
XX  
DT 12-FEB-1999 (first entry)  
XX Protein encoded by the consensus sequence of the PS108 gene.  
DE  
XX PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy.  
XX Homo sapiens.  
XX WO9850567-A1.  
XX 12-NOV-1998.  
XX  
XX 01-MAY-1998; 98WO-US08930.  
XX  
XX 02-MAY-1997; 97US-0850713.  
XX

PA (ABBO ) ABBOTT LAB.  
XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
XX WPI; 1999-034731/03.  
XX N-PSDB; AAV71181.  
XX New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
XX Claim 17; Pages 99-100; 122pp; English.  
XX The present sequence is encoded by the consensus sequence for a PS108  
CC gene. The sequence is derived from overlapping clones AAV71166-79. The  
CC clone sequences are PS108 gene-specific. They are used in the method  
CC of the invention. The specification describes a method for detecting the  
CC presence of a target PS108 polynucleotide in a test sample. The method  
CC comprises contacting the test sample with at least 1 PS108-specific  
CC polynucleotide or complement, and detecting the presence of the target  
CC PS108 polynucleotide. The products can be used for detecting,  
CC diagnosing, staging, monitoring, prognosticating, in vivo imaging,  
CC preventing or treating, or determining predisposition to diseases or  
CC conditions of the prostate such as benign prostatic hyperplasia (BPH),  
CC prostatitis, prostatic intraepithelial neoplasia (PIN) and cancer. In  
CC particular the products can be used in drug screening and gene therapy.  
XX Sequence 255 AA;  
SQ

Query Match 89.7%; Score 131; DB 20; Length 255;  
Best Local Similarity 100.0%; Pred. No. 3e-11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GLYQGVPRAPGTEARRHYDEGVR 27  
Db 1 glyqgvpraepgtearrhydegvr 24

RESULT 8  
AAB29268  
ID AAB29268 standard; Protein; 255 AA.  
XX  
AC AAB29268;  
XX  
DT 07-FEB-2001 (first entry)  
XX Human prostate-related PS108 partial protein sequence SEQ ID NO: 36.  
DE Human; prostate cancer; PS108; antibody; tumour; metastasis.  
XX Homo sapiens.  
XX US6130043-A.  
XX 10-OCT-2000.  
XX  
XX 01-MAY-1998; 98US-0071710.  
XX  
XX 02-MAY-1997; 97US-0850713.  
XX  
XX (ABBO ) ABBOTT LAB.  
XX Cohen M, Colpitts TL, Friedman PN, Gordon J, Granados EN;  
PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;  
PI Kratochvil JD, Russell JC, Hodges SC;  
XX WPI; 2000-655655/63.  
XX Methods for detecting target prostate-specific polynucleotides or  
PT diseases of the prostate (e.g. prostate cancer), comprising detecting



PT the presence of any of PS108 nucleic acid sequences in a test sample -  
XX  
PS Example 1; Column 85-88; 55pp; English.  
XX  
CC The present invention is related to a number of partial coding and  
CC protein sequences for the human prostate tissue protein PS108. These  
CC sequences can be used in the diagnosis and prognosis of prostate  
CC diseases, particularly prostate cancer. They can also be used to produce  
CC antibodies which can be used in treatment. The present sequence is one  
CC of the PS108 partial protein sequences.  
XX  
SQ Sequence 255 AA;  
  
Query Match 89.7%; Score 131; DB 21; Length 255;  
Best Local Similarity 100.0%; Pred. No. 3e-11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 4 GLYQGVPRAEPTGTEARRHYDEGVR 27  
|||||  
Db 1 glyqgvpraepgtearrhydegvr 24  
  
RESULT 9  
AAY94977  
ID AAY94977 standard; Protein; 742 AA.  
XX  
AC AAY94977;  
XX  
DT 16-JUN-2000 (first entry)  
XX  
DE Human secreted protein clone as180\_1 protein sequence SEQ ID NO:160.  
XX  
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;  
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;  
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;  
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;  
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;  
KW connective tissue disease; multiple sclerosis; erythematosis;  
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;  
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;  
KW insulin dependent diabetes mellitus; graft-versus-host-disease;  
KW autoimmune inflammatory eye disease; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200009552-A1.  
XX  
PD 24-FEB-2000.  
XX  
PF 13-AUG-1999; 99WO-US18298.  
XX  
PR 14-AUG-1998; 98US-0096622.  
PR 17-AUG-1998; 98US-0096815.  
PR 04-SEP-1998; 98US-0099229.  
PR 23-OCT-1998; 98US-0105368.  
PR 08-JAN-1999; 99US-0115234.  
PR 12-FEB-1999; 99US-0119931.  
PR 18-FEB-1999; 99US-0120575.  
PR 30-APR-1999; 99US-0132020.  
PR 11-AUG-1999; 99US-0096622.  
XX  
PA (GEMY ) GENETICS INST INC.  
XX  
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
PI Wong GG, Clark HF, Fechtel K;  
XX  
DR WPI: 2000-205979/18.  
XX  
PT New polynucleotides encoding secreted proteins, which may have e.g.  
PT nutritional, chemokine, immune stimulating or suppressing,  
PT hematopoiesis regulating, tissue growth, activin/inhibin

PT antiinflammatory or tumor inhibition activity -  
XX  
PS Claim 169; Page 614-616; 641pp; English.  
XX  
CC AAA16618 to AAA16697 encode the human secreted proteins given in  
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,  
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,  
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,  
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,  
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are  
CC predicted to have biological activities which would make them suitable  
CC for treating, preventing or ameliorating medical conditions in humans  
CC and animals. The polynucleotides can be used as markers for tissues in  
CC which the protein is preferentially expressed, as molecular weight  
CC markers on Southern gels, and as chromosome markers or tags to identify  
CC chromosomes or to map gene positions. The proteins can be used in the  
CC treatment of immune deficiencies and disorders, such as severe combined  
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other  
CC infections. These infections include human immunodeficiency virus (HIV),  
CC hepatitis, herpesviruses, mycobacteria, leishmania spp., malaria and  
CC candidiasis. The proteins can be used to treat autoimmune disorders such  
CC as connective tissue disease, multiple sclerosis, systemic lupus  
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,  
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent  
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and  
CC autoimmune inflammatory eye disease. The proteins can also be used to  
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent  
CC probes for the human secreted proteins from the present invention.  
XX  
SQ Sequence 742 AA;  
  
Query Match 43.2%; Score 63; DB 21; Length 742;  
Best Local Similarity 51.7%; Pred. NO. 0.66;  
Matches 15; Conservative 2; Mismatches 8; Indels 4; Gaps 1;  
  
QY 3 EGLYQGVPRAEPTGTEARRHYD----EGVR 27  
|||  
Db 95 egeyqgipraesgkggermadgaplagvr 123  
  
RESULT 10  
AAB09155  
ID AAB09155 standard; Protein; 27 AA.  
XX  
AC AAB09155;  
XX  
DT 30-AUG-2000 (first entry)  
XX  
DE Hepatitis GB virus protein sequence SEQ ID NO:277.  
XX  
KW Hepatitis GB virus; HGBV; diagnosis; therapeutic; immunogenic;  
KW infection; detection; characterisation; hepatitis.  
XX  
OS Hepatitis GB virus.  
XX  
PN US6051374-A.  
XX  
PD 18-APR-2000.  
XX  
PF 07-JUN-1995; 95US-0488445.  
XX  
PR 14-FEB-1994; 94US-0196030.  
PR 13-MAY-1994; 94US-0242654.  
PR 29-JUL-1994; 94US-0283314.  
PR 23-NOV-1994; 94US-0344185.  
PR 23-NOV-1994; 94US-0344190.  
PR 30-JAN-1995; 95US-0377557.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Dawson GJ, Leary TP, Muerhoff AS, Pilot-Matias TJ, Buijk SL;  
PI Mushahwar IK, Simons JN, Desai SM, Erker JC, Schlauder GC;



PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.

```
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 32.9%; Score 48; DB 21; Length 113;
Best Local Similarity 54.5%; Pred. No. 12;
Matches 12; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 3 EGLY---QGVPRAPGTEARR 20
Db 53 egllrrdraraeprgrqarr 74

RESULT 13
AAG27261
ID AAG27261 standard; Protein; 123 AA.
AC AAG27261;
XX 17-OCT-2000 (first entry)
DT
DE
DE
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS zea mays subsp. mays.
XX
PN EP1033405-A2.
PD
PD
PD 06-SEP-2000.
PF
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
```



PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 32.9%; Score 48; DB 21; Length 123;  
Best Local Similarity 54.5%; Pred. No. 14;  
Matches 12; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 3 EGLY---QGVPRAEFGTEARR 20  
||| : | |||| :|||  
Db 53 eglrrdravaraeprqarr 74

RESULT 14  
AAY41125  
ID AAY41125 standard; protein; 563 AA.  
XX  
AC AAY41125;  
XX

DT 17-JAN-2000 (first entry)  
DE Wheat sucrose transport protein (clone wlmk1.pk0002.e11).  
XX  
KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;  
KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;  
KW carbohydrate transport; plant cell.  
XX

OS Triticum aestivum.

XX WO9953068-A2.

XX 21-OCT-1999.

PF 07-APR-1999; 99WO-US07562.

PR 09-APR-1998; 98US-0081148.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Allen SM, Hitz WD, Rafalski JA;

XX WPI; 1999-620432/53.

DR N-PSDB; AA223135.

XX New sucrose transport proteins from plants, useful for controlling

PT carbohydrate transport and distribution in plant cells -

XX Claim 5; Page 56-57; 64pp; English.

CC The invention provides nucleic acid sequences (AA223124-223135) encoding  
CC sucrose transport proteins (AAY41114-Y41125) derived from corn, rice,  
CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can  
CC be recombinantly expressed by standard recombinant methodology. The  
CC invention facilitates studies on carbohydrate metabolism and function in  
CC plants, provides genetic tools for the manipulation of these  
CC biosynthetic pathways, and provides a means to control carbohydrate  
CC transport and distribution in plant cells.

XX Sequence 563 AA;

Query Match 32.9%; Score 48; DB 20; Length 563;  
Best Local Similarity 37.0%; Pred. No. 72;  
Matches 10; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 1 VEGELYQGVPRAEFGTEARRHYDEGVR 27

Db 360 mgrevyhgdpkg--naserkayddgvr 384

RESULT 15  
AAY53798  
ID AAY53798 standard; peptide; 90 AA.  
XX AC AAY53798;  
XX DT 22-FEB-2000 (first entry)  
XX DE Amino acids 145-234 of the mature human chromogranin A (CgA) protein.  
XX KW Human; chromogranin A; CgA; pheochromocytoma; epitope;  
XX KW intestinal carcinoma.  
XX OS Homo sapiens.  
XX XX WO9958980-A1.  
XX PN 18-NOV-1999.  
XX PD 12-MAY-1999; 99WO-FR01141.  
XX PF 14-MAY-1998; 98FR-0006101.  
XX PR (CISB-) CIS BIO INT.  
XX PA Degorce F, Bellanger L, Aunis D;  
XX PI WPI; 2000-039179/03.  
XX DR New immunological assay of human chromogranin A and kit, useful for  
XX PT diagnosis and monitoring pheochromocytomas and intestinal carcinomas -  
XX PS Claim 1; Page -: 54pp; French.  
XX XX AAY53798-Y53800 represent fragments of the mature human chromogranin A  
XX CC (CgA) protein. The fragments contain CgA epitopes recognised by the  
XX CC antibodies of the invention. The CgA protein was isolated from the  
XX CC glands of patients suffering from pheochromocytomas. The specification  
XX CC describes an immunological assay of CgA which comprises reacting at  
XX CC least one monoclonal or polyclonal antibody which specifically binds  
XX CC with an epitope in the human CgA N-terminal end. The epitopes are  
XX CC located between amino acids 145-234 of the CgA protein. Hybridomas  
XX CC producing the antibodies are produced from mice immunized with the CgA  
XX CC protein. The N-terminal sequences of the epitopes recognised by the  
XX CC antibodies of the invention are given in AAY53789-96. The assay is  
XX CC useful for diagnosing and monitoring pathologies such as  
XX CC pheochromocytomas and intestinal carcinomas. The assay allows the  
XX CC positive discrimination between the serum of patients with  
XX CC pheochromocytomas or intestinal carcinomas and normal serum.  
XX CC note: this sequence does not appear in the specification; it was created  
XX CC using information provided.

XX SQ Sequence 90 AA;  
Query Match 32.2%; Score 47; DB 21; Length 90;  
Best Local Similarity 42.9%; Pred. No. 14;  
Matches 12; Conservative 4; Mismatches 6; Indels 6; Gaps 1;  
OY 3 EGLYQGVV-----RAEPGTEARRHYDE 24  
Db 42 eglsqglvdrekglsaepegwgakreeee 69  
||| ||: |||| :||: ||  
RESULT 16  
AAY53797  
ID AAY53797 standard; protein; 439 AA.  
XX AC AAY53797;  
XX PN

XX 22-FEB-2000 (first entry)  
XX DE Amino acid sequence of the mature human chromogranin A (CgA) protein.  
XX KW Human; chromogranin A; CgA; pheochromocytoma; epitope;  
XX KW intestinal carcinoma.  
XX OS Homo sapiens.  
XX XX WO9958980-A1.  
XX PN 18-NOV-1999.  
XX PD 12-MAY-1999; 99WO-FR01141.  
XX PF 14-MAY-1998; 98FR-0006101.  
XX PR (CISB-) CIS BIO INT.  
XX PA Degorce F, Bellanger L, Aunis D;  
XX PI WPI; 2000-039179/03.  
XX DR New immunological assay of human chromogranin A and kit, useful for  
XX PT diagnosis and monitoring pheochromocytomas and intestinal carcinomas -  
XX PS Disclosure; Fig 1A-B; 54pp; French.  
XX XX The present sequence represents the mature human chromogranin A (CgA)  
XX CC protein. The protein was isolated from the glands of patients suffering  
XX CC from pheochromocytomas. The specification describes an immunological  
XX CC assay of CgA which comprises reacting at least one monoclonal or  
XX CC polyclonal antibody which specifically binds with an epitope in the  
XX CC human CgA N-terminal end. The epitopes are located between amino acids  
XX CC 145-234 of the CgA protein. Hybridomas producing the antibodies are  
XX CC produced from mice immunized with the CgA protein. The N-terminal  
XX CC sequences of the epitopes recognised by the antibodies of the invention  
XX CC are given in AAY53789-96. The assay is useful for diagnosing and  
XX CC monitoring pathologies such as pheochromocytomas and intestinal  
XX CC carcinomas. The assay allows the positive discrimination between the  
XX CC serum of patients with pheochromocytomas or intestinal carcinomas and  
XX CC normal serum.

XX SQ Sequence 439 AA;  
Query Match 32.2%; Score 47; DB 21; Length 439;  
Best Local Similarity 42.9%; Pred. No. 76;  
Matches 12; Conservative 4; Mismatches 6; Indels 6; Gaps 1;  
OY 3 EGLYQGVV-----RAEPGTEARRHYDE 24  
Db 186 eglsqglvdrekglsaepegwgakreeee 213  
||| ||: |||| :||: ||  
RESULT 17  
AAG14138  
ID AAG14138 standard; Protein; 913 AA.  
XX AC AAG14138;  
XX DT 17-OCT-2000 (first entry)  
XX XX Arabidopsis thaliana protein fragment SEQ ID NO: 13887.  
XX DE Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
XX XX Arabidopsis thaliana.  
XX OS EP1033405-A2.  
XX PN



PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999;	99US-0123548.
PR	23-SEP-1999;	99US-0155486.	PR	23-MAR-1999;	99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999;	99US-0126264.
PR	28-SEP-1999;	99US-0156438.	PR	29-MAR-1999;	99US-0126785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999;	99US-0127462.
PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999;	99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159330.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159331.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159637.	PR	05-MAY-1999;	99US-0132485.
PR	18-OCT-1999;	99US-0159584.	PR	06-MAY-1999;	99US-0132486.
PR	21-OCT-1999;	99US-0160741.	PR	06-MAY-1999;	99US-0132487.
PR	21-OCT-1999;	99US-0160767.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160770.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160815.	PR	14-MAY-1999;	99US-0134219.
PR	22-OCT-1999;	99US-0160980.	PR	14-MAY-1999;	99US-0134221.
PR	22-OCT-1999;	99US-0160981.	PR	14-MAY-1999;	99US-0134370.
PR	25-OCT-1999;	99US-0160989.	PR	18-MAY-1999;	99US-0134768.
PR	25-OCT-1999;	99US-0161404.	PR	19-MAY-1999;	99US-0134941.
PR	25-OCT-1999;	99US-0161405.	PR	20-MAY-1999;	99US-0135124.
PR	25-OCT-1999;	99US-0161406.	PR	21-MAY-1999;	99US-0135353.
PR	26-OCT-1999;	99US-0161359.	PR	24-MAY-1999;	99US-0135629.
PR	26-OCT-1999;	99US-0161360.	PR	25-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161361.	PR	27-MAY-1999;	99US-0136392.
PR	28-OCT-1999;	99US-0161920.	PR	28-MAY-1999;	99US-0136782.
PR	28-OCT-1999;	99US-0161922.	PR	01-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161992.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161993.	PR	04-JUN-1999;	99US-0137529.
PR	29-OCT-1999;	99US-0162142.	PR	07-JUN-1999;	99US-0137724.
Query Match 32.2%; Score 47; DB 21; Length 913; Pred. NO. 1.7e+02; Mismatches 3; Indels 0; Gaps 0;					
Best Local Similarity 50.0%;					
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;					
QY	7	QGVPRAEPTGTEARRHYDE 24	PR	10-JUN-1999;	99US-0138094.
Db	849	rgprkrpeterkrifde 866	PR	10-JUN-1999;	99US-0138540.
RESULT 18					
AAG37416					
ID	AAG37416 standard; Protein; 916 AA.				
XX	AAG37416;				
AC	AAG37416;				
XX	18-OCT-2000 (first entry)				
DT	Arabidopsis thaliana protein fragment SEQ ID NO: 46002.				
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 46002.				
XX	Protein identification; signal transduction pathway; metabolic pathway;				
KW	hybridisation assay; genetic mapping; gene expression control; promoter;				
KW	termination sequence.				
XX	Arabidopsis thaliana.				
OS	EP1033405-A2.				
XX	06-SEP-2000.				
PN	25-FEB-2000; 2000EP-0301439.				
XX	25-FEB-1999; 99US-0121825.				
XX	05-MAR-1999; 99US-0123180.				
PR					



PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 32.2%; Score 47; DB 21; Length 916;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 QGVPRAEPTGTEARRHYDE 24  
: | | | | | | | : | |  
Db 852 rgrprkrpeterkrlfde 869

RESULT 19

AAG14137.  
ID AAG14137 standard; Protein; 994 AA.

XX AC AAG14137;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 13886.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 27-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 32.2%; Score 47; DB 21; Length 994;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 QGVPRAEPGTEARRHYDE 24  
:| || | || :| ||  
Db 930 rgrprkrpeterkrlfde 947

RESULT 20  
AAG37415  
ID AAG37415 standard; Protein; 997 AA.  
XX AC AAG37415;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 46001.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 31-AUG-1999; 99US-0151930.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 32.2%; Score 47; DB 21; Length 997;  
Best Local Similarity 50.0%; Pred. NO. 1.9e+02;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
  
QY 7 QGVPRAEPTGTEARRHYDE 24  
Db 933 rgrprkrpeterkrifde 950  
  
RESULT 21  
AAG141136  
ID AAG141136 standard; Protein; 1098 AA.  
XX AAG141136;  
AC AAG141136;  
XX 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 13885.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.



PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.

```
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 32.2%; Score 47; DB 21; Length 1098;
Best Local Similarity 50.0%; Pred. NO. 2.1e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps -0;

OY 7 QGVPRAEPCTEARRHYDE 24
Db 1034 rgrprkrpeterkrifide 1051

:| || | || :| ||
:| || | || :| ||

RESULT 22
AAG37414
ID AAG37414 standard; Protein; 1101 AA.
XX
AC AAG37414;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46000.
DE Arabidopsis thaliana
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
```

```
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
```

PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 32.2%; Score 47; DB 21; Length 1101;  
Best Local Similarity 50.0%; Pred. No. 2.1e+02;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
  
QY 7 QGVPRAEPTGTEARRHYDE 24  
: | | | | | | | : | |  
Db 1037 rgrprkrpeterkrifde 1054  
  
RESULT 23  
AAW42074  
ID AAW42074 standard; Protein; 140 AA.  
XX  
AC AAW42074;  
XX  
DT 07-JUL-1998 (first entry)  
XX  
DE The amino acid sequence of the C195\_1 protein.  
XX  
KW C195\_1 protein; human PBMC cDNA library; nutritional activity;  
KW peripheral blood mononuclear cell; nutritional activity;  
KW cytokine activity; cell proliferation/differentiation activity;  
KW homology; EST; expressed sequence tag.  
XX  
OS Homo sapiens.  
XX  
PN WO9748801-A2.  
XX  
PD 24-DEC-1997.  
XX  
PF 16-JUN-1997; 97WO-US10501.  
XX  
PR 17-JUN-1996; 96US-0664596.  
XX  
PA (GEM) GENETICS INST INC.  
XX  
PI Bowman M, Evans C, Jacobs K, Lavallie ER, McCoy JM;  
PI Merberg D, Racie LA, Spaulding V, Treacy M;  
XX  
DR WPI; 1998-063142/06.  
DR N-PSDB; AAV09268.  
XX  
PT Poly:nucleotide(s) and proteins obtained from human PBMC, dendritic  
PT cell, adult brain, foetal brain and adult testes cDNA libraries -  
PT used in research, detection and therapy of, e.g. cytokine and cell  
PT proliferation or differentiation  
XX  
PS Claim 1; Pages 47-48; 78pp; English.  
XX  
CC This is the amino acid sequence of the C195\_1 protein which was  
CC isolated from a human PBMC cDNA library. It can be used in research,  
CC detection and therapy, as they may have nutritional activity, cytokine  
CC and cell proliferation/differentiation activity. A search of the  
CC sequence encoding this protein against the Genbank database  
CC demonstrated that this sequence has at least some homology with two ESTs  
CC identified as "ygl1906.r1 Homo sapiens cDNA clone 31937 5'" and  
CC "ym15f12.r1 Homo sapiens cDNA clone 48025 5'".  
XX  
SQ Sequence 140 AA;

Query Match 31.8%; Score 46.5; DB 19; Length 140;  
Best Local Similarity 47.4%; Pred. No. 26;  
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;  
  
QY 5 LYQGVPRAEPTGTEARRHYD 23  
: : | | | | | | : | | |  
Db 83 vwsghpaempgk-rhhfd 100







```

XX SR-p70; monkey; transcription factor; p53; tumour suppressor gene;
KW homology; differential splicing; diagnosis; cancer; neuroblastoma;
KW gene therapy; apoptosis.
XX Cercopithecus aethiops.
XX OS
XX PN WO9728186-A1.
XX PD
XX PF 07-AUG-1997.
XX PR 03-FEB-1997; 97WO-FR00214.
XX PR 02-FEB-1996; 96FR-0001309.
XX PA (SNFI ) SANOFI SA.
XX PI Caput D, Ferrara P, Kaghad AM;
XX DR WPI; 1997-402550/37.
XX DR N-PSDB; AAV01497.
XX PT New polypeptide(s) encoded by the SR-p70 tumour suppressor gene -
XX PT and related nucleic acid, useful for diagnosis and treatment of
XX PT tumours
XX PS
XX Claim 1; Fig 5; 136pp; French.
XX This is the amino acid sequence of the protein SR-p70b from monkey
XX cells. SR-p70 are transcription factors which may control the activity
XX of p53-regulated genes, and are expressed by tumour suppressor genes
XX related to the p53 gene family. The gene sequence was isolated from a
XX cDNA library by sequencing the inserts and comparing to sequence
XX databases. The SR-p70b gene sequence was isolated simultaneously with the
XX protein. The SR-p70b gene sequence was isolated simultaneously with the
XX SR-p70a sequence (AAV01496) from the library and is created by
XX differential splicing of the SR-p70 mRNA sequence. The sequences can be
XX used in the diagnosis and monitoring of cancer, especially neuroblastoma.
XX The nucleic acid sequences and corresponding antisense sequences, are
XX also useful in gene therapy, e.g. to regulate apoptosis.
XX
XX Sequence 499 AA;
XX
XX Query Match 31.5%; Score 46; DB 18; Length 499;
XX Best Local Similarity 56.2%; Pred. No. 1.2e+02;
XX Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
XX
XX QY 9 VPRAEPGTEARRHYDE 24
XX II : III II
XX Db 337 vpalpgvgvkkrrhgde 352
XX
XX RESULT 29
XX AAR72328
XX ID AAR72328 standard; Protein; 529 AA.
XX AC AAR72328;
XX DT 14-OCT-1995 (first entry)
XX DE Laccase Rslac3.
XX KW Rslac3; laccase; lignin; lignosulfonate; polymerization;
XX depolymerization.
XX OS Rhizoctonia solani.
XX PN WO9507988-A.
XX PD 23-MAR-1995.
XX PF 13-SEP-1994; 94WO-US10264.
XX

```

```

XX 17-SEP-1993; 93US-0122230.
XX 17-SEP-1993; 93US-0122827.
XX 03-DEC-1993; 93US-0162827.
XX 22-DEC-1993; 93US-0172331.
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO-NORDISK AS.
XX Christensen BE, Schneider P, Wahleithner JA;
XX WPI; 1995-131356/17.
XX DR N-PSDB; AAQ86526.
XX New neutral Rhizoctonia laccase(s) and corresp. nucleic acids -
XX PT are used industrially for polymerising lignin, lignosulphate(s)
XX PT and phenolic cpds. and for oxidising dyes.
XX PS
XX Claim 6; Page 44-46; 78pp; English.
XX A R. solani RS22 (IMI CC 358730) cDNA library was screened on
XX ABTS. The DNA insert of a plasmid from an isolated clone was
XX sequenced to reveal a novel gene, Rslac3 (given in AAQ86526),
XX encoding a laccase (AAR72328) optimally active at pH 6.0-8.5.
XX Recombinant laccase was expressed in E. coli.
XX
XX Sequence 529 AA;
XX
XX Query Match 31.5%; Score 46; DB 16; Length 529;
XX Best Local Similarity 66.7%; Pred. No. 1.3e+02;
XX Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 6 YQGVPRAPGTE 17
XX II : III II
XX Db 314 yegapnaeptte 325
XX
XX RESULT 30
XX AAW16301
XX ID AAW16301 standard; Protein; 529 AA.
XX AC AAW16301;
XX DT 07-AUG-1997 (first entry)
XX DE Rhizoctonia solani laccase isozyme 4.
XX KW Blue copper oxidase; laccase; enzyme engineering;
XX protein engineering; lignin depolymerisation; dye oxidation.
XX OS Rhizoctonia solani.
XX FH Key
XX FT Binding-site 427
XX FT Binding-site /note= "Type I copper site ligand"
XX FT Binding-site 470 /note= "possible Type I copper site ligand"
XX FT Binding-site 480 /note= "Type I copper site ligand"
XX FT Binding-site 485 /note= "Type I copper site ligand"
XX FT Peptide 217..226 /note= "preferred site for mutation (Claim 30),
XX corresponds to W163 region of ascorbate
XX oxidase"
XX FT Peptide 303..312 /note= "preferred site for mutation (Claim 29),
XX corresponds to R285 region of ascorbate
XX oxidase"
XX FT Peptide 356..371 /note= "preferred site for mutation (Claim 28),
XX corresponds to Q353 and W362 region of
XX FT

```

FT ascorbate oxidase"  
FT 416..429  
FT /note= "preferred site for mutation (Claim 26),  
FT contains Cu-ligating His"  
FT  
FT Peptide 465..472  
FT /note= "preferred site for mutation (Claim 31)"  
FT Peptide 466..470  
FT /note= "preferred site for mutation (Claim 18)"  
FT Peptide 474..494  
FT /note= "preferred site for mutation (Claim 27),  
FT contains Cu-ligating His and Cys"  
FT  
FT Misc-difference 379  
FT /note= "translated residue from gene sequence is  
FT Ile"  
FT  
FT Misc-difference 415  
FT /note= "translated residue from gene sequence is  
FT Ile"  
FT  
XX  
PN WO9709431-A1.  
XX  
XX 13-MAR-1997.  
PD  
XX  
XX 03-SEP-1996; 96WO-US14087.  
PF  
XX  
XX 01-SEP-1995; 95US-0003142.  
PR  
XX  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
PA  
XX  
PI Berka RM, Wahleithner JA, Xu F, Berka R;  
XX  
XX WPI; 1997-192906/17.  
DR  
XX N-PSDB; AAT63317.  
XX  
PT New mutant blue copper oxidase enzymes - having different specific  
PT activities to wild-type enzymes, used for e.g. lignin  
PT (de)polymerisation or oxidation of dyes  
XX  
PS Disclosure; Fig 6A-D; 48pp; English.  
XX  
CC Rhizoctonia solani laccase isozyme 4 (AAW16301) can be engineered to  
CC alter e.g. its specific activity or pH-activity profile or to  
CC improve expression yields. The mutant laccase is prepd. by site-  
CC directed mutagenesis of the rsl4 gene (AAT63317) and expression in  
CC host cells. The mutation may be a deletion, insertion or pref. a  
CC substn. of one or more amino acids at a location no greater than  
CC 12 Angstroms, pref. no more than 2.5 Angstroms, from the Type I  
CC copper site. The mutant enzyme can be used e.g. for the  
CC polymerisation or depolymerisation of lignins, oxidation of dyes,  
CC or polymerisation of phenolic or aniline cpds. present in liquids,  
CC e.g. for the treatment of fruit juices.  
XX  
SQ Sequence 529 AA;  
  
Query Match 31.5%; Score 46; DB 18; Length 529;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 6 YQGVPRAEPTGE 17  
I: I I I I I I I  
Db 314 yegapnaeptte 325  
  
RESULT 31  
AAW76310  
ID AAW76310 standard; protein; 529 AA.  
XX  
AC AAW76310;  
XX  
DT 08-JAN-1999 . (first entry)  
XX  
DE Rhizoctonia solani (I) laccase protein.  
XX

KW Laccase; variant; oxidisation; dye transfer inhibition; bleaching;  
KW denim; lignin modification; paper strengthening; phenol polymerisation;  
KW hair dye; waste water treatment.  
XX  
OS Rhizoctonia solani.  
XX  
XX WO9838287-A1.  
PN  
XX  
PD 03-SEP-1998.  
XX  
XX 23-FEB-1998; 98WO-DK00070.  
PF  
XX  
PR 28-FEB-1997; 97DK-0000222.  
XX  
XX (NOVO ) NOVO-NORDISK AS.  
PA  
XX Svendsen A, Xu F;  
PI  
XX WPI; 1998-495393/42.  
DR  
XX  
XX New variants of Coprinus and related laccases with increased  
PT oxidation potential - or altered pH optimum, or mediator or  
PT oxygen-hydroxide ion pathways, useful for oxidation, for inhibiting  
PT dye transfer and in bleaching textiles, especially as detergent  
PT additive  
XX  
PS Disclosure; Pages 124-125; 147pp; English.  
XX  
XX The present sequence represents a laccase protein. The specification  
CC describes active laccase variants (see AAW76282, AAW76296-99 and  
CC AAW76316-17) having increased oxidation potential, altered pH optimum,  
CC altered mediator and/or altered oxygen/hydroxide ion pathway. The  
CC laccase variants are used specifically to oxidise substrates, to  
CC inhibit dye transfer, and for bleaching textiles, specifically denim.  
CC They can also be used for lignin modification, strengthening paper,  
CC polymerisation of phenols, dyeing of hair and textiles and waste  
CC water treatment.  
XX  
SQ Sequence 529 AA;  
  
Query Match 31.5%; Score 46; DB 19; Length 529;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 6 YQGVPRAEPTGE 17  
I: I I I I I I I  
Db 314 yegapnaeptte 325  
  
RESULT 32  
AAW60878  
ID AAW60878 standard; Protein; 529 AA.  
XX  
AC AAW60878;  
XX  
DT 09-NOV-1998 (first entry)  
XX  
DE Rhizoctonia solani (I) laccase.  
XX  
KW Coprinus laccase-like enzyme; enzyme engineering; enzyme stability;  
KW detergent; bleaching.  
XX  
OS Rhizoctonia solani.  
XX  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 126 /note= "variant may have Ala, Val, Leu, Ile, Pro,  
FT Phe, Met, Gly, Ser, Thr, Cys, Tyr, Asn,  
FT Gln, Asp, Glu, Lys, Arg or His (preferably  
FT Phe or His) at this position"  
FT  
FT Misc-difference 135 /note= "variant may have Ala, Val, Leu, Ile, Pro,  
FT

FT Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,  
FT Gln, Asp, Glu, Lys, Arg or His (preferably  
FT Phe) at this position"  
FT  
FT Misc-difference 127 /note=  
FT "variant may have Ala, Val, Leu, Ile, Pro,  
FT Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,  
FT Gln, Asp, Glu, Lys, Arg or His (preferably  
FT Phe) at this position"  
FT  
FT Misc-difference 171 /note=  
FT "variant may have Ala, Val, Leu, Ile, Pro,  
FT Phe, Trp, Gly, Ser, Thr, Cys, Met, Asn,  
FT Gln, Asp, Glu, Lys, Arg or His (preferably  
FT Phe) at this position"  
FT  
FT Misc-difference 76 /note=  
FT "variant may have Ala, Val, Leu, Ile, Pro,  
FT Phe, Trp, Gly, Ser, Thr, Cys, Tyr, Asn,  
FT Gln, Asp, Glu, Lys, Arg or His (preferably  
FT Phe, val, Ile, Leu or Gln) at this  
FT position"  
FT  
XX WO9827198-A1.  
XX PN  
XX XX  
XX 25-JUN-1998.  
XX PD  
XX 16-DEC-1997; 97WO-DK00571.  
XX PF  
XX 08-SEP-1997; 97DK-0001021.  
XX PR  
XX 19-DEC-1996; 96DK-0001449.  
XX XX  
XX (NOVO ) NOVO-NORDISK AS.  
XX PA  
XX Cherry JR, Pedersen AH, Rasmussen G, Schneider P;  
XX Svendsen A;  
XX WPI; 1998-362768/31.  
XX DR  
XX  
XX New laccase variants with improved stability - having amino acid  
XX changes based on Coprinus laccase structure, used for e.g.  
XX oxidation, dye transfer inhibition or bleaching  
XX  
XX Claim 13; 147-148; 168pp; English.

This is a laccase enzyme of Rhizoctonia solani. The invention relates to the design of new variants of Coprinus-like laccases (see AAW60874-79, AAW60925 and AAW62501-03); the R. solani laccase shows 64.8% homology to C. cinereus laccase. The modifications are based on the previously unknown three-dimensional structure of C. cinereus laccase. Amino acid residues identified as being important to protein stability are identified and altered to improve stability. The variants are typically obtained by mutagenesis of laccase DNA and expression in a host cell. Variants are preferably modified within 15 (especially 10 or 5) Angstrom of a copper ion in the three-dimensional structure of the laccase. For R. solani laccase variants, preferred substitutions are one or more of W128F/H, Y137F, Y129F, Y171F or M78F/V/L/Q. The stabilised laccase variants can be used in detergent additives, for dye transfer inhibition in detergents, in bleaching of textiles (in particular denim), for lignin modification, paper strengthening, phenol polymerisation, hair dyeing and in waste water treatment.

Query Match 31.5%; Score 46; DB 19; Length 529;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Sequence 529 AA;

QY 6 YQGVPRAEPTGE 17  
DB 314 yegapnaeptte 325

RESULT 33

AAAY50728  
ID AAY50728 standard; Protein; 529 AA.  
XX  
XX AAY50728;  
AC  
XX 08-FEB-2000 (first entry)  
DT  
XX  
XX R. solani laccase isozyme 4 protein.  
DE  
XX  
XX Laccase; mutant; lignin polymerization; Kraft; liganosulfate; lignin;  
KW depolymerization; phenol; fruit juice; dye.  
XX Rhizoctonia solani.  
OS  
XX US5972670-A.  
PN  
XX 26-OCT-1999.  
XX  
XX 09-JAN-1998; 98US-0005397.  
XX  
XX 30-AUG-1996; 96US-0706037.  
XX  
XX (FENG/) FENG X.  
PA (BERK/) BERKA R M.  
PA (WAHL/) WAHLEITHNER J A.  
XX  
XX Berka RM, Feng X, Wahleithner JA;  
PI WPI; 2000-021805/02.  
XX N-PSDB; AAZ24235.  
DR  
XX Mutant Rhizoctonia laccase enzymes useful for the polymerization and  
PT depolymerization of lignin -  
PT  
XX Claim 3; Column 31-34; 43pp; English.

This invention describes novel mutant Rhizoctonia solani laccase enzymes modified to exhibit altered characteristics (e.g. pH activity profile) relative to the wild-type enzyme. The mutant enzymes have a number of commercial and industrial applications. For example, they may be used for the polymerization of lignin (both Kraft and liganosulfates) in solution to produce high molecular weight lignin. They may also be used for the in situ depolymerization of lignin in Kraft pulp to produce low molecular weight lignin. They may also be used for the polymerization of phenolic or aniline compounds in liquids (for example apple juice may be treated with the enzyme to accelerate precipitation of the phenolic compounds in the juice to stabilize it). Additionally, it may be used to decolorize dyes and dye precursors by oxidation. The mutant laccase enzymes exhibit different characteristics to those of the wild-type enzyme. In particular, they may exhibit an altered pH activity profile relative to the wild-type oxidase or altered specific activities. Additionally, they may be produced in large quantities. This sequence represents the Rhizoctonia solani laccase isozyme 4 which is described in the method of the invention.

Sequence 529 AA;

Query Match 31.5%; Score 46; DB 21; Length 529;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAEPTGE 17  
DB 314 yegapnaeptte 325

RESULT 34

AAAB25538  
ID AAB25538 standard; Protein; 541 AA.  
XX  
XX AAB25538;  
AC  
XX



DT 27-NOV-2000 (first entry)  
XX Eucalyptus grandis cell signalling involved protein SEQ ID NO:913.  
DE  
XX  
KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;  
KW environmental change; development; cell proliferation; differentiation;  
KW elongation; survival; disease resistance; nutrient metabolism.  
XX  
OS Eucalyptus grandis.  
XX  
PN WO200042171-A1.  
XX  
PD 20-JUL-2000.  
XX  
PF 11-JAN-2000; 2000WO-US00724.  
XX  
PR 12-JAN-1999; 99US-0228986.  
PR 01-NOV-1999; 99US-0162866.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
PI Strabala TJ, Nieuwenhuizen NJ;  
XX  
DR WPI; 2000-476052/41.  
XX  
PT Isolated polynucleotide encoding a polypeptide involved in cell  
PT signaling used for generating transgenic plants with modified responses  
PT to external signals -  
XX  
PS Claim 3; Page 475-476; 527pp; English.  
XX  
CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
CC are involved in cell signalling. The polynucleotide and protein  
CC sequences can be used to modify the response of plant cells to external  
CC signals e.g. environmental changes or pathogens during the growth and  
CC development of a plant. They can be used to modify cell proliferation,  
CC differentiation, elongation and survival, resistance to disease and  
CC nutrient metabolism. Examples of modifications which can be produced are  
CC altered fruit ripening and senescence of leaves and flowers e.g. to  
CC delay senescence and prolong the life of cut flowers or enhance  
CC senescence of reproductive organs to engineer sterile plants. Other  
CC modifications can be used to delay senescence in selected cell types or  
CC organs providing fruit and vegetables which have a longer shelf life  
CC between harvest and consumption, or to decrease branching frequency in  
CC forest tree species giving long stretches of valuable knot-free clear  
CC wood which can be used in solid timber furniture and veneers.  
XX  
SQ Sequence 541 AA;  
  
Query Match 31.5%; Score 46; DB 21; Length 541;  
Best Local Similarity 40.9%; Pred. No. 1.3e+02;  
Matches 9; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
  
QY 5 LYQGVPRAEPTGTEARRHYDEGV 26  
|||:| |||:  
Db 237 lyggipetisdmealehldvgi 258  
  
RESULT 35  
AAW86348  
ID AAW86348 standard; Protein; 566 AA.  
XX  
AC AAW86348;  
XX  
DT 15-MAR-1999 (first entry)  
XX  
DE Human multiple endocrine receptor neoplasia type 1 protein.  
XX  
KW Human; multiple endocrine neoplasia type 1 protein; MEND-1;

KW tumorigenesis; cancer; multiple endocrine neoplasia; gene therapy;  
KW detection; diagnosis; drug screening.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 137  
FT /label= unknown  
FT /note= "encoded by YGG"  
FT Misc-difference 161  
FT /label= unknown  
FT /note= "encoded by NNT"  
FT Misc-difference 274  
FT /label= unknown  
FT /note= "encoded by RGG"  
FT Misc-difference 312  
FT /label= unknown  
FT /note= "encoded by SCC"  
FT Misc-difference 417  
FT /label= unknown  
FT /note= "encoded by TNG"  
FT Misc-difference 421  
FT /label= unknown  
FT /note= "encoded by NGC"  
FT Misc-difference 424  
FT /label= unknown  
FT /note= "encoded by GAN"  
FT Misc-difference 428  
FT /label= unknown  
FT /note= "encoded by CNG"  
FT Misc-difference 435  
FT /label= unknown  
FT /note= "encoded by ANC"  
FT Misc-difference 443  
FT /label= unknown  
FT /note= "encoded by NCC"  
XX  
PN WO9854324-A1.  
XX  
PD 03-DEC-1998.  
XX  
PF 29-MAY-1998; 98WO-US10957.  
XX  
PR 29-MAY-1997; 97US-0865337.  
XX  
PA (INCY-) INCYTE PHARM INC.  
PI Au-Young J, Covitz PA, Murry LE, Tang YT;  
XX  
DR WPI; 1999-059839/05.  
DR N-PSDB; AAV80659.  
XX  
PT New isolated multiple endocrine neoplasia type 1 protein - used to  
PT develop products for the diagnosis, treatment and prevention of  
PT cancers and multiple endocrine neoplasia.  
XX  
PS Claim 1; Fig 1; 67pp; English.  
XX  
CC The present sequence is human multiple endocrine neoplasia type 1  
CC protein (MEND-1). MEND-1 plays a role in multiple endocrine neoplasia  
CC when one or both normal genetic copies of MEND-1 are mutated and no  
CC longer able to suppress tumorigenesis. MEND-1 can be used for treating  
CC or preventing cancers and multiple endocrine neoplasia. MEND-1  
CC polynucleotides can also be used for gene therapy. Products from the  
CC present invention can also be used for detection, diagnosis and drug  
CC screening.  
XX  
SQ Sequence 566 AA;  
  
Query Match 31.5%; Score 46; DB 20; Length 566;  
Best Local Similarity 52.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEA 18  
|:| | | | | | |  
Db 463 gggavsgpprkppgtva 479

RESULT 36  
AAW29749  
ID AAW29749 standard; Protein; 610 AA.  
XX AC  
XX AAW29749;  
DT 11-JAN-1999 (first entry)  
XX DE  
DE Tumor suppressor protein menin.  
XX KW  
KW Familial multiple endocrine neoplasia type 1; FMEN1; MEN1;  
KW menin; tumour suppressor gene; cancer; marker; diagnosis;  
KW gene therapy; human.  
XX OS  
XX Homo sapiens.  
XX PN  
PN WO9839439-A1.  
XX DE  
PD 11-SEP-1998.  
XX PF  
PF 04-MAR-1998; 98WO-US04258.  
XX PR  
PR 05-MAR-1997; 97US-0040269.  
XX PA  
(USSH ) US DEPT HEALTH & HUMAN SERVICES.

PI Agarwal SK, Burns AL, Chandrasekharappa SC, Collins FS;  
PI Debelenko LV, Emmert-Buck MR, Guru SC, Liotta LA;  
PI Lubensky IA, Manickam P, Marx SJ, Spiegel AM;  
DR WPI; 1998-506360/43.  
DR N-PSDB; AAV57415-16.

Protein and gene associated with multiple endocrine neoplasia type 1  
- useful in gene therapy and to diagnose sufferers of, and those  
susceptible to, this condition by detecting protein absence or gene  
mutation(s)

Claim 11; Page 56-57; 75pp; English.

This is the amino acid sequence of menin, a 67.5 kDa polypeptide  
encoded by a novel human tumour suppressor gene, MEN1 (see AAV57416),  
that is associated with multiple endocrine neoplasia type 1. Menin  
exhibits no apparent amino acid sequence similarities to previously  
known proteins. The lack of a functional menin polypeptide, either  
by absence of the protein, its alteration and/or associated  
mutations in the MEN1 gene, have been identified in individuals  
with familial multiple endocrine neoplasia type 1 (FMEN1) and  
suffering from multiple endocrine neoplasia type 1. The  
identification of MEN1 provides a new window into the mechanism  
of endocrine tumorigenesis, facilitates accurate early diagnosis of  
MEN1 associated cancers, and provides preclinical identification of  
individuals with the FMEN1 syndrome, i.e. cancer-free individuals  
that are at high risk of acquiring MEN1 associated tumours. It  
also provides a basis for gene therapy. Menin can be expressed in  
eukaryotic or prokaryotic host cells and used to identify menin in  
specific reagents (e.g. antibodies) useful for detecting menin in  
humans cells or tissues, especially from patients suspected of  
being at risk from multiple endocrine neoplasia type 1.

Sequence 610 AA;

Query Match 31.5%; Score 46; DB 19; Length 610;  
Best Local Similarity 52.9%; Pred. NO. 1.5e+02;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEA 18  
|:| | | | | | |  
Db 507 gggavsgpprkppgtva 523

RESULT 37  
AAB40554  
ID AAB40554 standard; Protein; 619 AA.  
XX AC  
XX AAB40554;  
DT 08-FEB-2001 (first entry)  
XX DE  
DE Human ORFX ORF318 polypeptide sequence SEQ ID NO:636.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
anticongulant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
hypotensive; dermatological; immunosuppressive; antiinflammatory;  
antiviral; antibacterial; antifungal; antineumatic; antithyroid;  
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
neurodegenerative disorder; osteoarthritis; graft vs host disease;  
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
cholesterol ester storage; systemic lupus erythematosus; infection;  
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
bone damage; cartilage damage; antiinflammatory disease; coagulation;  
thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.

N-PSDB; AAC74763.

Novel nucleic acids and peptides derived from open reading frame X,  
useful for treating e.g. cancers, proliferative disorders,  
neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 760-761; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
which represent the human ORFX open reading frames 1 to 3161. The ORFX  
sequences have activities such as: cytostatic; hepatotropic; vulnary;  
antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
osteopathic; anticongulant; antiarthritic; immunosuppressant;  
immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
antidiabetic; hypotensive; dermatological; immunosuppressive;  
antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
antithyroid; and antianaemic. The sequences can be used for determining  
the presence of or predisposition to, or preventing or treating  
pathological conditions associated with an ORFX-associated disorder. The  
nucleic acids can be used to express ORFX proteins in gene therapy  
vectors. The proteins and nucleic acids may be used to treat cancers,  
proliferative disorders, neurodegenerative disorders, osteoarthritis,  
graft vs host disease, cardiovascular disease, diabetes mellitus,  
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 619 AA;  
  
Query Match 31.5%; Score 46; DB 21; Length 619;  
Best Local Similarity 29.6%; Pred. No. 1.6e+02;  
Matches 8; Conservative 10; Mismatches 9; Indels 0; Gaps 0;  
  
QY 1 VGEGLYQGVPRAPGTEARRHYDEGVR 27  
:|: :|:| |:| :|:| |:  
Db 384 mgqvifegdpkapsnstawqaynagvk 410  
  
RESULT 38  
AAW36182  
ID AAW36182 standard; Protein; 637 AA.  
XX  
AC AAW36182;  
XX  
DT 27-APR-1998 (first entry)  
XX  
DE Monkey p53 tumour suppressor-related protein SR-p70a.  
XX  
KW SR-p70; monkey; transcription factor; p53; tumour suppressor gene;  
KW homology; differential splicing; diagnosis; cancer; neuroblastoma;  
KW gene therapy; apoptosis.  
XX  
OS Cercopithecus aethiops.  
XX  
PN WO9728186-A1.  
XX  
PD 07-AUG-1997.  
XX  
PF 03-FEB-1997; 97WO-FR00214.  
XX  
PR 02-FEB-1996; 96FR-0001309.  
XX  
PA (SNFI ) SANOFI SA.  
XX  
PI Caput D, Ferrara P, Kaghad AM;  
XX  
DR WPI; 1997-402550/37.  
DR N-PSDB; AAV01496.  
XX  
PT New polypeptide(s) encoded by the SR-p70 tumour suppressor gene -  
PT and related nucleic acid, useful for diagnosis and treatment of  
PT tumours  
XX  
PS Claim 1; Fig 4; 136pp; French.  
XX  
CC This is the amino acid sequence of the protein SR-p70a from monkey  
CC cells. SR-p70 are transcription factors which may control the activity  
CC of p53-regulated genes, and are expressed by tumour suppressor genes  
CC related to the p53 gene family. The gene sequence was isolated from a  
CC cDNA library by sequencing the inserts and comparing to sequence  
CC databases. The protein sequence contains regions of homology to the p53  
CC protein. A second sequence (SR-p70b) was also isolated from the library  
CC and was caused by differential splicing of the sequence (see AAV01497).  
CC The sequences can be used in the diagnosis and monitoring of cancer,  
CC especially neuroblastoma. The nucleic acid sequences and corresponding  
CC antisense sequences, are also useful in gene therapy, e.g. to regulate  
CC apoptosis.  
XX  
SQ Sequence 637 AA;

Query Match 31.5%; Score 46; DB 18; Length 637;  
Best Local Similarity 56.2%; Pred. No. 1.6e+02;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 9 VPRAEPGTEARRHYDE 24  
|| ||: ||| ||  
Db 337 vpalgpgvkkrrhgde 352  
  
RESULT 39  
AAR88634  
ID AAR88634 standard; Protein; 1088 AA.  
XX  
AC AAR88634;  
XX  
DT 27-JUN-1996 (first entry)  
XX  
DE Plasmid pASK75 open reading frame (a) translation.  
XX  
KW Tetracycline; resistance; TetR gene; repressor protein; TetA gene;  
KW prokaryotic expression vector; beta-lactamase promoter; bla gene;  
KW circular; cyclic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1..1088 /note= "each X corresponds to a termination codon"  
FT Peptide 89..98 /label= Strep\_tag  
FT /note= "streptavidin-binding peptide"  
XX  
PN WO9532295-A1.  
XX  
PD 30-NOV-1995.  
XX  
PF 17-MAY-1995; 95WO-EP01862.  
XX  
PR 19-MAY-1994; 94DE-4417598.  
XX  
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX  
PI Skerra A, Wardenberg C;  
XX  
DR WPI; 1996-020588/02.  
DR N-PSDB; AAT11192.  
XX  
PT Prokaryotic vector for regulated prodn. of heterologous protein gene  
PT - controlled by promoter repressed by tetracycline repressor  
PT protein, prodn. of which is controlled by independent promoter not  
PT regulated by this repressor  
XX  
PS Disclosure; Fig 1a; 50pp; German.  
XX  
CC Plasmid pASK75 is a specifically claimed example of a prokaryotic  
CC expression vector comprising the tetracycline promoter/operator (P/O)  
CC region and the tetracycline repressor (tetR) gene. The tetR gene is  
CC under the control of the beta-lactamase (bla) promoter. Expression of  
CC heterologous genes inserted downstream of the tet P/O is controlled  
CC by the activity of the tet repressor. Plasmid pASK75 was derived  
CC from pASK60-Strep (Schmidt and Skerra, Protein. Eng. 6 (1993), 109-  
CC 122) and further contains the OmpA leader peptide coding sequence, a  
CC polylinker and a streptavidin-binding peptide strep-tag coding the  
CC region from the parent plasmid. In addition, pASK75 contains the  
CC ColEI origin of replication, the bla gene, intergene regions of  
CC filamentous phage f1 and a lipoprotein terminator.  
CC The specification includes the translated sequences from all three  
CC reading frames of pASK75; the Strep tag coding region is in reading  
CC frame (a).  
XX  
SQ Sequence 1088 AA;

Query Match 31.5%; Score 46; DB 17; Length 1088;  
Best Local Similarity 50.0%; Pred. No. 2.9e+02;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;





OS Pseudomonas sp.  
 XX DE19850242-A1.  
 PN 04-MAY-2000.  
 PD 31-OCT-1998; 98DE-1050242.  
 XX 31-OCT-1998; 98DE-1050242.  
 PF (HAAR ) HAARMANN & REIMER GMBH.  
 XX Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;  
 PI WPI; 2000-340642/30.  
 XX N-PSDB; AAA39731.  
 DR Transformed organisms, used for production of vanillin and other  
 XX methoxyphenols, have altered catabolism of eugenol or ferulic acid  
 PT Disclosure; Fig 2r; 80pp; German.  
 XX This invention describes novel transformed and/or mutated uni- or  
 CC multi-cellular organisms (A) in which enzymes (I) involved in catabolism  
 CC of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way  
 CC that the organism accumulates the intermediates coniferyl alcohol (CAL),  
 CC coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).  
 CC INDEPENDENT CLAIMS are also included for the following: (a) gene  
 CC structures in which sequences that encode one or more of the enzymes  
 CC CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA  
 CC hydratase-aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase  
 CC have been altered and/or inactivated; (b) vector containing at least one  
 CC gene structure of (a); (c) biotechnical production of organic compounds  
 CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)  
 CC production method of (A). (A) are used to prepare organic compounds,  
 CC preferably alcohols, aldehydes and organic acids, especially CAL, CA2,  
 CC FA, V and/or VA, most particularly V, a well-known natural flavouring.  
 CC The inexpensive starting material eugenol can be converted to V in a  
 CC single step. This sequence represents the Pseudomonas sp. HR199 vanillin  
 CC dehydrogenase, vdh-delta which is described in the method of the  
 CC invention.  
 XX SQ Sequence 411 AA;  
 Query Match 30.8%; Score 45; DB 21; Length 411;  
 Best Local Similarity 47.1%; Pred. NO. 1.4e+02;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 4 GLYQGVPRAPGTEARR 20  
 III III: I :I  
 Db 81 glysgvpkvspalgvqr 97  
 RESULT 43  
 AAY87980  
 ID AAY87980 standard; Protein; 586 AA.  
 XX AAY87980;  
 XX 21-SEP-2000 (first entry)  
 DT Pseudomonas sp. HR199 vdh-omega-Gm protein.  
 XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;  
 KW vanillic acid; flavouring; vanillin dehydrogenase; vdh;  
 KW vdh-omega-Gm.  
 XX Pseudomonas sp.  
 OS DE19850242-A1.  
 PN 04-MAY-2000.  
 PD

XX 31-OCT-1998; 98DE-1050242.  
 PF 31-OCT-1998; 98DE-1050242.  
 XX (HAAR ) HAARMANN & REIMER GMBH.  
 PA Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;  
 XX WPI; 2000-340642/30.  
 DR N-PSDB; AAA39730.  
 XX Transformed organisms, used for production of vanillin and other  
 PT methoxyphenols, have altered catabolism of eugenol or ferulic acid  
 XX Disclosure; Fig 2q; 80pp; German.  
 XX This invention describes novel transformed and/or mutated uni- or  
 CC multi-cellular organisms (A) in which enzymes (I) involved in catabolism  
 CC of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way  
 CC that the organism accumulates the intermediates coniferyl alcohol (CAL),  
 CC coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).  
 CC INDEPENDENT CLAIMS are also included for the following: (a) gene  
 CC structures in which sequences that encode one or more of the enzymes  
 CC CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA  
 CC hydratase-aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase  
 CC have been altered and/or inactivated; (b) vector containing at least one  
 CC gene structure of (a); (c) biotechnical production of organic compounds  
 CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)  
 CC production method of (A). (A) are used to prepare organic compounds,  
 CC preferably alcohols, aldehydes and organic acids, especially CAL, CA2,  
 CC FA, V and/or VA, most particularly V, a well-known natural flavouring.  
 CC The inexpensive starting material eugenol can be converted to V in a  
 CC single step. This sequence represents the Pseudomonas sp. HR199 vanillin  
 CC dehydrogenase, vdh-omega-Gm which is described in the method of the  
 CC invention.  
 XX SQ Sequence 586 AA;  
 Query Match 30.8%; Score 45; DB 21; Length 586;  
 Best Local Similarity 47.1%; Pred. NO. 2e+02;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 4 GLYQGVPRAPGTEARR 20  
 III III: I :I  
 Db 81 glysgvpkvspalgvqr 97  
 RESULT 44  
 AAY87979  
 ID AAY87979 standard; Protein; 673 AA.  
 XX AAY87979;  
 AC 21-SEP-2000 (first entry)  
 DT Pseudomonas sp. HR199 vdh-omega-Km protein.  
 XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;  
 KW vanillic acid; flavouring; vanillin dehydrogenase; vdh;  
 KW vdh-omega-Km.  
 XX Pseudomonas sp.  
 OS DE19850242-A1.  
 PN 04-MAY-2000.  
 PD 31-OCT-1998; 98DE-1050242.  
 XX 31-OCT-1998; 98DE-1050242.  
 PR

PA (HAAR ) HAARMANN & REIMER GMBH.  
XX Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;  
PI WPI; 2000-340642/30.  
XX N-PSDB; AAA87979.  
DR Transformed organisms, used for production of vanillin and other  
XX methoxyphenols, have altered catabolism of eugenol or ferulic acid  
PT Disclosure; Fig 2p; 80pp; German.  
XX  
XX This invention describes novel transformed and/or mutated uni- or  
CC multi-cellular organisms (A) in which enzymes (I) involved in catabolism  
CC of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way  
CC that the organism accumulates the intermediates conferyl alcohol (CA1),  
CC conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).  
CC INDEPENDENT CLAIMS are also included for the following: (a) gene  
CC structures in which sequences that encode one or more of the enzymes  
CC CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA  
CC hydratase-aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase  
CC have been altered and/or inactivated; (b) vector containing at least one  
CC gene structure of (a); (c) biotechnical production of organic compounds  
CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)  
CC production method of (A). (A) are used to prepare organic compounds,  
CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,  
CC FA, V and/or VA, most particularly V, a well-known natural flavouring.  
CC The inexpensive starting material eugenol can be converted to V in a  
CC single step. This sequence represents the pseudomonas sp. HRL99 vanillin  
CC dehydrogenase, vdh-Omega-Km which is described in the method of the  
XX invention.  
XX  
XX Sequence 673 AA;  
SQ  
  
Query Match 30.8%; Score 45; DB 21; Length 673;  
Best Local Similarity 47.1%; Pred. NO. 2.4e+02;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
  
QY 4 GLYQGVPRAEPTGTEARR 20  
Db 81 glysgvpsvpalgvqr 97  
  
RESULT 45  
AABL19929  
ID AABL19929 standard; Protein; 157 AA.  
XX  
XX AABL19929;  
XX  
XX 19-MAR-2001 (first entry)  
XX Human oxidoreductase OXRD-4.  
XX  
XX OXRD-4; human; oxidoreductase; cell proliferation;  
KW neurological disease; smooth muscle disease; autoimmune disease;  
KW inflammation; antiproliferative; neuroprotective;  
KW immunosuppressive; antiinflammatory; therapy; diagnosis.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Modified-site 2 /note= "O-phosphorylated"  
FT Modified-site 29 /note= "O-phosphorylated"  
FT Modified-site 105 /note= "O-phosphorylated"  
FT Domain 41..92 /note= "nitrogenase Mo/Fe protein domain"  
FT  
XX WO200071679-A2.  
XX

PD 30-NOV-2000.  
XX 19-MAY-2000; 2000WO-US13879.  
XX  
XX 20-MAY-1999; 99US-0135049.  
PR 27-MAY-1999; 99US-0136740.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX Yue H, Tang YT, Baughn MR, Lu DAM;  
PI WPI; 2001-025146/03.  
XX N-PSDB; AAA89188.  
DR  
DR New human oxidoreductase proteins useful for diagnosing, treating or  
XX preventing proliferative, neurological, genetic, smooth muscle,  
XX autoimmune or inflammatory disorders associated with abnormal  
XX expression of oxidoreductase proteins -  
XX Claim 1(a); Page 84; 95pp; English.  
XX  
XX The present sequence is that of human oxidoreductase OXRD-4, as  
XX deduced from a cDNA clone (see AAA89188) isolated from a foetal  
XX kidney cDNA library. The protein shows homology to Bacillus subtilis  
XX nitrite reductase subunit, and includes a nitrogenase Mo/Fe protein  
XX domain. OXRD-4 is expressed in developmental, cardiovascular,  
XX haematopoietic/immune and reproductive tissue, and may be involved  
XX in cell proliferation and inflammation. The invention provides  
XX OXRD-1 to -8 polypeptides (see AABL9926-33) and polynucleotides (see  
XX AAA89185-92). It also provides methods for using these polypeptides  
XX and polynucleotides for diagnosing, treating or preventing disorders  
XX associated with expression of OXRD, especially cell proliferative,  
XX neurological, genetic, smooth muscle, and autoimmune/inflammatory  
XX disorders. A vector encoding OXRD or its fragments is useful for  
XX somatic or germline gene therapy for treating such disorders.  
XX The proteins can also be used to screen for agonists and  
XX antagonists useful for treating these conditions, while antibodies  
XX that bind to OXRD may be used for diagnosis or in assays to monitor  
XX patient treatment.  
XX  
XX Sequence 157 AA;  
SQ  
  
Query Match 30.5%; Score 44.5; DB 22; Length 157;  
Best Local Similarity 41.4%; Pred. NO. 57;  
Matches 12; Conservative 3; Mismatches 5; Indels 9; Gaps 2;  
  
QY 2 GEGLYQGV----PRAEP-----GTEARRH 21  
Db 92 geglyqsinpkdpsakpkwscgikqrlh 120  
  
Search completed: June 28, 2001, 11:54:17  
Job time: 78 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:01 ; Search time 42.81 Seconds  
(without alignments)  
48.043 Million cell updates/sec

Title: US-09-439-313-566  
Perfect score: 146  
Sequence: 1 VGEGLYQGVRAEPGTEARRHYDEGVR 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	43.2	742	1 A43344	synaptic vesicle p
2	63	43.2	742	2 S27263	synaptic vesicle p
3	50.5	34.6	610	2 F83384	probable asparagin
4	49	33.6	420	2 T50585	probable membrane
5	49	33.6	2314	2 T28698	hypothetical prote
6	48	32.9	153	2 T37219	probable lipoprote
7	48	32.9	558	1 S03707	glutamate dehydrog
8	48	32.9	616	2 JQ1441	hypothetical 67K p
9	47.5	32.5	175	2 T09249	late embryonic abu
10	47.5	32.5	309	1 XNECV	branched-chain-ami
11	47.5	32.5	413	2 JC2520	beta-fructofuranos
12	47	32.2	127	2 A82594	Apag protein XF214
13	47	32.2	346	2 C75139	bifunctional short
14	47	32.2	457	1 A28468	chromogranin A pre
15	47	32.2	484	2 JE0341	R-SMAD protein, Sm
16	47	32.2	558	1 DEHUE	glutamate dehydrog
17	47	32.2	921	2 H84921	hypothetical prote
18	47	32.2	962	2 G86479	hypothetical prote
19	46.5	31.8	110	2 C75382	hypothetical prote
20	46.5	31.8	309	2 A86063	branched-chain ami
21	46	31.5	249	2 T09407	stationary-phase s
22	46	31.5	529	2 S68120	laccase (EC 1.10.3
23	46	31.5	558	1 A53719	glutamate dehydrog
24	46	31.5	583	1 B64816	ABC-type transport
25	46	31.5	583	2 A85588	hypothetical prote
26	46	31.5	592	2 E75032	carbon starvation
27	46	31.5	744	1 QQBEA7	UL69 protein - hum
28	46	31.5	1741	2 T15978	hypothetical prote
29	45.5	31.2	105	2 A43861	vir-repressed prot

30	45.5	31.2	308	1 A34082	branched-chain-ami
31	45.5	31.2	413	2 S47527	extracellular sucra
32	45.5	31.2	714	2 T00942	hypothetical prote
33	45.5	31.2	806	2 T01164	hypothetical prote
34	45.5	31.2	819	2 A84504	Mutator-like trans
35	45	30.8	38	2 B29502	carboxylesterase (
36	45	30.8	324	2 JC2395	Fas antigen precur
37	45	30.8	557	2 S38908	glutathione reduct
38	45	30.8	565	1 A65012	hypothetical prote
39	45	30.8	565	2 F85880	probable sensor pr
40	45	30.8	582	2 A84206	hypothetical prote
41	45	30.8	1460	1 EDBEIF	immediate-early pr
42	45	30.8	2124	2 A28452	proteoglycan core
43	44.5	30.5	142	2 E84396	hypothetical prote
44	44.5	30.5	235	2 E72783	hypothetical prote
45	44.5	30.5	413	2 E70661	probable PE protei
46	44	30.1	196	2 H72461	hypothetical prote
47	44	30.1	202	2 F72739	hypothetical prote
48	44	30.1	213	2 C82953	hypothetical prote
49	44	30.1	215	1 E72546	L-fucose-phospha
50	44	30.1	216	2 JC5039	adenylate kinase (
51	44	30.1	216	2 G84324	adenylate kinase (
52	44	30.1	217	1 KIPC	hypothetical prote
53	44	30.1	220	2 B84304	conserved hypothet
54	44	30.1	262	2 E75478	hypothetical prote
55	44	30.1	371	2 E72665	hypothetical prote
56	44	30.1	406	2 T47674	hypothetical prote
57	44	30.1	562	2 JC5182	serine C-palmitoyl
58	44	30.1	595	2 T49384	related to NRDL pr
59	44	30.1	730	2 T16455	hypothetical prote
60	44	30.1	809	1 C56279	carbon-monoxide de
61	44	30.1	926	2 H84415	DNA-directed RNA p
62	44	30.1	972	1 A33926	DNA-directed RNA p
63	44	30.1	1323	2 PN0568	connectin 3B - chi
64	44	30.1	7962	2 I38346	elastic titin - hu
65	43.5	29.8	152	2 JQ1244	coat protein 2 - b

ALIGNMENTS

RESULT 1  
A43344  
synaptic vesicle protein SV2 - rat  
N;Alternate names: transporter-like protein p87  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Sep-2000  
C;Accession: A43344; A43267; A58948  
R;Feany, M.B.; Lee, S.; Edwards, R.H.; Buckley, K.M.  
Cell 70, 861-867, 1992  
A;Title: The synaptic vesicle protein SV2 is a novel type of transmembrane transporte  
A;Reference number: A43344; MUID:92386605  
A;Accession: A43344  
A;Molecule type: mRNA  
A;Residues: 1-742 <FEA>  
A;Cross-references: GB:L01788; NID:g207090  
A;Experimental source: pheochromocytoma cell line PC12  
A;Note: sequence extracted from NCBI backbone (NCBIN:112840, NCBIP:112842)  
A;Note: this ORF is not annotated in GenBank entry RATS2A, release 109.0  
R;Bajjalieh, S.M.; Peterson, K.; Shinghal, R.; Scheller, R.H.  
Science 257, 1271-1273, 1992  
A;Title: SV2, a brain synaptic vesicle protein homologous to bacterial transporters.  
A;Reference number: A43267; MUID:92390722  
A;Accession: A43267  
A;Molecule type: mRNA  
A;Residues: 1-339, 'F', 341-742 <BAJ>  
A;Cross-references: GB:L05435; NID:g207091; PIDN:AAA42188.1; PID:g207092  
A;Experimental source: brain  
A;Note: sequence extracted from NCBI backbone (NCBIP:112489)  
R;Gingrich, J.A.; Andersen, P.H.; Tiberi, M.; El Mestikawy, S.; Jorgensen, P.N.; Frem  
FEBS Lett. 312, 115-122, 1992  
A;Title: Identification, characterization, and molecular cloning of a novel transport  
A;Reference number: S27263; MUID:93050176



QY 2 GEGLYQGVPRAEFGTEARRHYDEGVR 27  
| | | | : : | | : | | |

RESULT 6  
T37219  
probable lipoprotein - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C/Accession: T37219  
R/Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1998  
A/Reference number: Z21615  
A/Accession: T37219  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-153 <OLI>  
A/Cross-references: EMBL:AL031514; PIDN:CAA20597.1; GSPDB:GN00070; SCOEDB:SC2H4.03  
A/Experimental source: strain A3(2)



A;Reference number: A64720; MUID:97426617  
A;Accession: E65180  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-309 <BLAT>  
A;Cross-references: GB:AE000453; GB:U00096; NID:g2367276; PIDN:AAC77490.1; PID:g1790205;  
A;Experimental source: strain K-12, substrain MG1655  
R;Lawther, R.P.; Wek, R.C.; Lopes, J.M.; Pereira, R.; Taillon, B.E.; Hatfield, G.W.  
Nucleic Acids Res. 15, 2137-2155, 1987  
A;Title: The complete, nucleotide sequence of the ilvGMDA operon of Escherichia coli K-12  
A;Reference number: A26570; MUID:87174741  
A;Accession: C26570  
A;Molecule type: DNA  
A;Residues: 1-150,'A',152-309 <LA1>  
A;Cross-references: GB:X04890; NID:g288528; PIDN:CAA28575.1; PID:g288531  
R;Kuramitsu, S.; Ogawa, T.; Ogawa, H.; Kagamiyama, H.  
J. Biochem. 97, 993-999, 1985  
A;Title: Branched-chain, amino acid aminotransferase of Escherichia coli: nucleotide sequence  
A;Reference number: A91996; MUID:85289113  
A;Accession: A91996  
A;Molecule type: DNA  
A;Residues: 1-150,'A',152-309 <KUR>  
A;Cross-references: GB:X02413; NID:g41787; PIDN:CAA26262.1; PID:g41789  
R;Experimental source: strain K12  
R;Lawther, R.P.; Nichols, B.; Zurawski, G.; Hatfield, G.W.  
Nucleic Acids Res. 7, 2289-2301, 1979  
A;Title: The nucleotide sequence preceding and including the beginning of the ilvE gene  
A;Reference number: A93693; MUID:80101081  
A;Accession: A93693  
A;Molecule type: DNA  
A;Residues: 1-57,'V',59-61,'R',63-81 <LA2>  
A;Experimental source: strain K12  
R;Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.  
Science 257, 771-778, 1992  
A;Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 85.5  
A;Reference number: S30660; MUID:92358234  
A;Accession: S30668  
A;Molecule type: DNA  
A;Residues: 1-309 <DAN>  
A;Cross-references: EMBL:M87049; NID:g836656; PIDN:AAA67573.1; PID:g148177  
R;Inoue, K.; Kuramitsu, S.; Aki, K.; Watanabe, Y.; Takagi, T.; Nishigai, M.; Ikai, A.; K  
J. Biochem. 104, 777-784, 1988  
A;Title: Branched-chain amino acid aminotransferase of Escherichia coli: overproduction  
A;Reference number: A57546; MUID:89174510  
A;Accession: A57546  
A;Molecule type: protein  
A;Residues: 2-31;150,'A',152-165;308-309 <INO>  
A;Note: identification of N6-pyridoxal phosphate lysine  
C;Genetics:  
A;Gene: ilvE  
A;Map position: 85 min  
C;Function:  
A;Description: EC 2.6.1.42 [validated; MUID:85289113]; catalyzes the reversible transami  
A;Pathway: branched-chain amino acid biosynthesis  
C;Superfamily: branched-chain-amino-acid aminotransferase  
C;Keywords: aminotransferase; branched-chain amino acid biosynthesis; phosphoprotein; py  
F;2-309/Product: branched-chain-amino-acid aminotransferase #status experimental  
F;160/Binding site: pyridoxal phosphate (lys) (covalent) #status experimental

Query Match 32.5%; Score 47.5; DB 1; Length 309;  
Best Local Similarity 31.4%; Pred. No. 27;  
Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;

QY 3 EGLYQGVV-----RAEP-----GTEARRH-YDEGV 26  
| | | | |  
Db 134 EALEQCIDAMVSSWNRPNPTIPTAAKAGNYLSLLVGSARRHGYQEGI 184  
| | | | |

RESULT 11  
JC2520  
beta-fructofuranosidase (EC 3.2.1.26) - Zymomonas mobilis  
N;Alternate names: invertase

C;Species: Zymomonas mobilis  
C;Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 20-Jun-2000  
C;Accession: JC2520; PC2377; S50037  
R;Kyono, K.; Yanase, H.; Tonomura, K.; Kawasaki, H.; Sakai, T.  
BioSci. Biotechnol. Biochem. 59, 289-293, 1995  
A;Title: Cloning and characterization of Zymomonas mobilis genes encoding extracellular  
A;Reference number: JC2519; MUID:95218269  
A;Accession: JC2520  
A;Molecule type: DNA  
A;Residues: 1-413 <KYO>  
A;Cross-references: DDBJ:D17524; NID:g809529; PIDN:BAA04476.1; PID:g809531  
A;Accession: PC2377  
A;Molecule type: protein  
A;Residues: 1-21 <KY2>  
R;Rhee, S.K.  
submitted to the EMBL Data Library, March 1994  
A;Reference number: S50037  
A;Accession: S50037  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-104,'S',106-107,'T',109-413 <RHE>  
A;Cross-references: EMBL:L08094; NID:g468429; PIDN:AAA61488.1; PID:g468430  
C;Genetics:  
A;Gene: sucZE3  
C;Keywords: glycosidase; hydrolase

Query Match 32.5%; Score 47.5; DB 2; Length 413;  
Best Local Similarity 52.4%; Pred. No. 37;  
Matches 11; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 2 GEGLYQGVPRAPGTEARRHY 22  
| | | | |  
Db 326 GSGLVGGPASQP-TEAYAHY 345

RESULT 12  
AB2594  
A;Accession: AB2594  
A;Title: Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C;Accession: AB2594  
R;anonymus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID:20365717  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: A82594  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-127 <SIM>  
A;Cross-references: GB:AE004029; GB:AE003849; NID:g9107276; PIDN:AAF84948.1; GSPDB:GN  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XF2149  
C;Superfamily: apag protein

Query Match 32.2%; Score 47; DB 2; Length 127;





Db 331 GLLSNVNRNEVEQTRRHIGKGV 354

RESULT 16

DEHUE

Glutamate dehydrogenase (NAD(P)+) (EC 1.4.1.3) precursor - human

N:Alternate names: glutamic dehydrogenase

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1989 #sequence revision 31-Mar-1989 #text\_change 01-Dec-2000

C:Accession: A28208; A27697; A27881; A27481; S00958; C46125; A46125; S10769; A00

R:Biochem. Biophys. Res. Commun. 152, 1395-1400, 1988

R:Julliard, J.H.; Smith, E.L.

R:Biochem. Biophys. Res. Commun. 152, 1395-1400, 1988

R:Proc. Natl. Acad. Sci. U.S.A. 85, 3494-3498, 1988

A:Title: Isolation and characterization of cDNA clones encoding human liver glutamate de

A:Reference number: A94195; MUID:88217927

A:Accession: A28208

A:Molecule type: mRNA

A:Residues: 1-558 <MAV>

A:Cross-references: GB:J03248; NID:g183053; PIDN:AAA52523.1; PID:g183054

A:Experimental source: liver

R:Amuro, N.; Yamaura, M.; Goto, Y.; Okazaki, T.

R:Biochem. Biophys. Res. Commun. 152, 1395-1400, 1988

A:Title: Molecular cloning and nucleotide sequence of the cDNA for human liver glutamate

A:Reference number: A90143; MUID:88240360

A:Accession: A27697

A:Molecule type: mRNA

A:Residues: 1-558 <MAV>

A:Cross-references: GB:M20867; NID:g183059; PIDN:AAA52526.1; PID:g183060

A:Experimental source: liver

A:Note: the authors translated the codon AUC for residue 380 as Gly

R:Banner, C.; Silverman, S.; Thomas, J.W.; Lampel, K.A.; Vitkovic, L.; Huie, D.; Wenthol

R:J. Neurochem. 49, 246-252, 1987

A:Title: Isolation of a human brain cDNA for glutamate dehydrogenase.

A:Reference number: A92973; MUID:87224937

A:Accession: A27881

A:Molecule type: mRNA

A:Residues: 301-558 <BAN>

A:Experimental source: brain

R:Nakatani, Y.; Banner, C.; von Herrath, M.; Schneider, M.E.; Smith, H.H.; Freese, E.

R:Biochem. Biophys. Res. Commun. 149, 405-410, 1987

A:Title: Comparison of human brain and liver glutamate dehydrogenase cDNAs.

A:Reference number: A90138; MUID:88106451

A:Accession: A27481

A:Molecule type: mRNA

A:Residues: 131-558 <NAK1>

A:Cross-references: GB:X07674; GB:M18377; NID:g31706; PIDN:CAA30521.1; PID:g31707

A:Note: the complete sequence and translation are not shown; the complete submitted sequ

R:Nakatani, Y.; Schneider, M.; Banner, C.; Freese, E.

Nucleic Acids Res. 16, 6237, 1988

A:Title: Complete nucleotide sequence of human glutamate dehydrogenase cDNA.

A:Reference number: S00958; MUID:88289377

A:Accession: S00958

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-558 <NAK2>

A:Cross-references: GB:M37154; NID:g183057; PIDN:AAA52525.1; PID:g183058; EMBL:X07769; N

R:Michaelidis, T.M.; Tzimagiorgis, G.; Moschonas, N.K.; Papamatheakis, J.

Genomics 16, 150-160, 1993

A:Title: The human glutamate dehydrogenase gene family: gene organization and structural

A:Reference number: A46125; MUID:93252371

A:Accession: C46125

A:Molecule type: DNA

A:Residues: 1-1 <MIC1>

A:Cross-references: GB:S60495; NID:g300129; PIDN:AAD14931.1; PID:g4262783

A:Accession: A46125

A:Molecule type: DNA

A:Residues: 150-215 <MIC2>

A:Cross-references: GB:S60498; NID:g300124; PIDN:AAC60587.1; PID:g300125

A:Note: sequence extracted from NCBI backbone (NCBIN:131434, NCBIN:131436, NCBIN:131438,

A:Accession: B46125

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 420-558 <MIC3>

A:Note: this translation from Fig. 6 is not annotated in GenBank, release 113.0

R:Amuro, N.; Goto, Y.; Okazaki, T.

R:Biochim. Biophys. Acta 1049, 216-218, 1990

A:Title: Isolation and characterization of the two distinct genes for human glutamate

A:Reference number: S10769; MUID:90304226

A:Accession: S10769

A:Molecule type: DNA

A:Residues: 150-215 <AMU2>

A:Cross-references: EMBL:X53144; NID:g31700; EMBL:X53145; NID:g31701; EMBL:X53146; NI

A:Note: this translation is not annotated in the cross-referenced GenBank entries

R:Julliard, J.H.; Smith, E.L.

R:Biochem. Biophys. Res. Commun. 152, 1395-1400, 1988

A:Title: Partial amino acid sequence of the glutamate dehydrogenase of human liver an

A:Reference number: A92250; MUID:79151102

A:Contents: annotation; partial sequence

C:Genetics:

A:Gene: GDB:GLUD1; GLUD; GDH1

A:Cross-references: GDB:119994; OMIM:138130

A:Map position: 10q23.3-10q23.3

A:Introns: 149/1; 176/1; 194/3; 216/1; 247/3; 307/3; 353/3; 399/3; 426/3; 468/1; 498/

A:Note: the list of introns may be incomplete

C:Complex: homohexamer

C:Function:

A:Description: catalyzes the reversible reaction of L-glutamate, water and NAD (or NA

C:Superfamily: glutamate dehydrogenase (NAD(P)+)

C:Keywords: homohexamer; mitochondrion; NAD; NADP; oxidoreductase

F:1-53/Domain: transit peptide (mitochondrion) #status predicted <TNP>

F:54-558/Product: glutamate dehydrogenase (NAD(P)+) #status predicted <MAT>

F:84,183/Binding site: substrate (Lys) #status predicted

Query Match 32.2%; Score 47; DB 1; Length 558;

Best Local Similarity 44.4%; Pred. No. 60;

Matches 12; Conservative 1; Mismatches 10; Indels 4; Gaps 1;

QY 4 GLYQGVPRAE-----GTEARRHYDEGV 26

Db 31 GWARGQPAAPQPGALAAARRHYSEAV 57

RESULT 17

H84921

hypothetical protein At2g47980 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: H84921

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: H84921

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-921 <STO>

A:Cross-references: GB:AE002093; NID:g3738310; PIDN:AAC63652.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g47980

A:Map position: 2

Query Match 32.2%; Score 47; DB 2; Length 921;

Best Local Similarity 50.0%; Pred. No. 1e+02;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 QGVPRAEPGTEARRHYDE 24

Db 857 RGRPRKRKPETERKRLFDE 874

RESULT 18

G86479

hypothetical protein AAF79886.1 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: G86479  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719  
A;Accession: G86479  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-962 <STO>  
A;Cross-references: GB:AE005172; NID:g8778971; PIDN:AAF79886.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1

Query Match 32.2%; Score 47; DB 2; Length 962;  
Best Local Similarity 70.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGEGLYQGV 10  
:|:|:|:|:|  
Db 820 IGDGLYEGVP 829

RESULT 19  
C75382  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
C;Accession: C75382  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896

A;Accession: C75382  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-110 <WHI>  
A;Cross-references: GB:AE001998; GB:AE000513; NID:g6459302; PIDN:AAF11112.1; PID:g645931  
A;Experimental source: strain R1  
C;Genetics:

A;Gene: DR1539  
A;Map position: 1  
C;Superfamily: Deinococcus radiodurans hypothetical protein DR1539

Query Match 31.8%; Score 46.5; DB 2; Length 110;  
Best Local Similarity 46.4%; Pred. No. 13;  
Matches 13; Conservative 2; Mismatches 10; Indels 3; Gaps 1;

QY 3 EGLYQGVPAEPG---TEARRHYDEGVR 27  
| : |||||  
Db 75 EKVAGVDRAKAGIHNAEAHASVDEGHR 102

RESULT 20  
A86063  
branched-chain amino-acid aminotransferase [imported] - Escherichia coli (strain O157:H7)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: A86063  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: A86063  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-309 <STO>  
A;Cross-references: GB:AE005174; NID:gl2518628; PIDN:AAG58965.1; GSPDB:GN00145; UWGP:  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: ilvE  
C;Superfamily: branched-chain-amino-acid aminotransferase

Query Match 31.8%; Score 46.5; DB 2; Length 309;  
Best Local Similarity 31.4%; Pred. No. 38;  
Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;

QY 3 EGLYQGV-----RAEP-----GTEARRH-YDEGV 26  
| | | | |  
Db 134 EALEQIDAMVSSWNRAAPTPTAAKAGGNLSSLLVGSEARRHGYQEGI 184

RESULT 21  
T09407  
stationary-phase survival protein sure - Haemophilus influenzae (strain Rd)  
C;Species: Haemophilus influenzae  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 23-Jul-1999  
C;Accession: T09407  
R;White, O.; Clayton, R.A.; Kerlavage, A.R.; Fleischmann, R.D.; Peterson, J.; Hickey,  
submitted to the EMBL Data Library, May 1998  
A;Reference number: Z16667  
A;Accession: T09407  
A;Status: preliminary; translated from GB/EMBL/DBBJ  
A;Molecule type: DNA  
A;Residues: 1-249 <WHI>  
A;Cross-references: EMBL:U32753; NID:gl573701; PID:g3212202  
C;Genetics:  
A;Gene: HI0702  
C;Superfamily: stationary-phase survival protein Sure

Query Match 31.5%; Score 46; DB 2; Length 249;  
Best Local Similarity 32.0%; Pred. No. 36;  
Matches 8; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 3 EGLYQGVPAEPGTEARRHYDEGVR 27  
| | : | : | : | : |  
Db 111 EGRHLGLPAITAVSLDGRQHYETAAR 135

RESULT 22  
S68120  
laccase (EC 1.10.3.2) 4 precursor - Rhizoctonia solani  
C;Species: Rhizoctonia solani  
C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000  
C;Accession: S68120; S77976; S72564  
R;Wahleithner, J.A.; Xu, F.; Brown, K.M.; Brown, S.H.; Golightly, E.J.; Halkier, T.;  
Curr. Genet. 29, 395-403, 1996  
A;Title: The identification and characterization of four laccases from the plant path  
A;Reference number: S68117; MUID:96171523  
A;Accession: S68120  
A;Molecule type: DNA  
A;Residues: 1-529 <WAH1>  
A;Accession: S77976  
A;Molecule type: protein  
A;Residues: 20-53;132-180;203-217;378-396;440-465;489-513;520-529 <WAH2>  
R;Wahleithner, J.A.  
submitted to the EMBL Data Library, September 1995  
A;Reference number: S72564  
A;Accession: S72564  
A;Molecule type: DNA  
A;Residues: 1-417, 'GAQHHR', 424-469, 'A', 471-483, 'CT', 484, 'R', 486, 'GSH', 490, 'S', 492, 'PR







extracellular sucrose - Zymomonas mobilis  
C:Species: Zymomonas mobilis  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Mar-1999  
C:Accession: S47527  
R:Song, K.B.; Lee, S.K.; Joo, H.K.; Rhee, S.K.  
Biochim. Biophys. Acta 1219, 163-166, 1994  
A:Title: Nucleotide and derived amino acid sequences of an extracellular sucrose gene (1  
A:Reference number: S47527; MUID:94368848  
A:Accession: S47527  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <SON>  
A:Cross-references: EMBL:L08094

Query Match 31.2%; Score 45.5; DB 2; Length 413;  
Best Local Similarity 52.4%; Pred. No. 72;  
Matches 11; Conservative 2; Mismatches 7; Indels 1; Gaps 1;  
QY 2 GEGLYQGVPRAPGTEARRHY 22  
| | | | | : | : | | | | |  
Db 326 GYGLVLGGPASQP-TEAYAHY 345

RESULT 32  
T00942  
hypothetical protein T3F12.3 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 31-Mar-2001  
C:Accession: T00942  
R:Gnoj, L.; Huang, E.N.; Habermann, K.; Hameed, A.; Hasegawa, A.; Jensen, K.; Schutz, K.  
submitted to the EMBL Data Library, October 1997  
A:Description: Arabidopsis thaliana BAC T3F12 from chromosome IV.  
A:Reference number: Z14210  
A:Accession: T00942  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-714 <GNO>  
A:Cross-references: EMBL:AC002983; NID:g2443899; PID:g2565002  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 359/3; 457/1; 531/1; 564/3  
A:Note: T3F12.3  
C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 31.2%; Score 45.5; DB 2; Length 714;  
Best Local Similarity 52.2%; Pred. No. 1.3e+02;  
Matches 12; Conservative 1; Mismatches 9; Indels 1; Gaps 1;  
QY 5 LYQGVPRAPGTEARRHYDEGVR 27  
| | : | | | | | | | | | |  
Db 391 LYM-IRRANPGTVARLQIDESRR 412

RESULT 33  
T01164  
hypothetical protein F7N22.10 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 02-Mar-2001  
C:Accession: T01164  
R:Dante, M.  
submitted to the EMBL Data Library, April 1998  
A:Description: The sequence of A. thaliana F7N22.  
A:Reference number: Z14250  
A:Accession: T01164  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-806 <DAN>  
A:Cross-references: EMBL:AF058825; NID:g3047060; PID:g3047068  
A:Experimental source: cultivar Columbia

C:Genetics:  
A:Map position: 4  
A:Introns: 380/3; 675/2  
C:Note: F7N22.10  
C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7  
Query Match 31.2%; Score 45.5; DB 2; Length 806;  
Best Local Similarity 52.2%; Pred. No. 1.4e+02;  
Matches 12; Conservative 1; Mismatches 9; Indels 1; Gaps 1;  
QY 5 LYQGVPRAPGTEARRHYDEGVR 27  
| | : | | | | | | | | | |  
Db 412 LYM-IRRANPGTVARLQIDESGR 433

RESULT 34  
A84504  
Mutator-like transposase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Mar-2001  
C:Accession: A84504  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: A84504  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-819 <STO>  
A:Cross-references: GB:AE002093; NID:g4850412; PIDN:AAD31079.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g12720  
A:Map position: 2  
C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 31.2%; Score 45.5; DB 2; Length 819;  
Best Local Similarity 52.2%; Pred. No. 1.5e+02;  
Matches 12; Conservative 1; Mismatches 9; Indels 1; Gaps 1;  
QY 5 LYQGVPRAPGTEARRHYDEGVR 27  
| | : | | | | | | | | | |  
Db 407 LYM-IRRANPGTVARLQIDESGR 428

RESULT 35  
B29502  
carboxylesterase (EC 3.1.1.1) 5 - fruit fly (Drosophila mojavensis) (fragment)  
N:Alternate names: esterase-5  
C:Species: Drosophila mojavensis  
C:Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 11-Jan-2000  
C:Accession: B29502  
R:Pen, J.; Van Beeumen, J.; Beintema, J.J.  
Biochem. J. 278, 691-699, 1986  
A:Title: Structural comparison of two esterases from Drosophila mojavensis isolated b  
A:Reference number: A90340  
A:Accession: B29502  
A:Molecule type: protein  
A:Residues: 1-38 <PEN>  
C:Genetics:  
A:Gene: FlyBase:Dmoj/Est-5  
A:Cross-references: FlyBase:FBgn0012569  
C:Superfamily: cholinesterase; cholinesterase homology  
C:Keywords: carboxylic ester hydrolase

Query Match 30.8%; Score 45; DB 2; Length 38;  
Best Local Similarity 64.3%; Pred. No. 7.1;  
Matches 9; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 3 EGLY--QGVPRAP 14  
||||| :|:|||||  
Db 22 EGYEAEIGIPRAEP 35

RESULT 36  
JC2395  
Fas antigen precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 05-Nov-1999  
C;Accession: JC2395; PC2246  
R;Kimura, K.; Wakatsuki, T.; Yamamoto, M.  
Biochem. Biophys. Res. Commun. 198, 666-674, 1994  
A;Title: A variant mRNA species encoding a truncated form of Fas antigen in the rat liver  
A;Reference number: JC2395; MUID:94128114  
A;Accession: JC2395  
A;Molecule type: mRNA  
A;Residues: 1-324 <KIM>  
A;Cross-references: DDBJ:D26112; NID:g468486; PIDN:BAA05108.1; PID:d1005650; PID:g468487  
A;Experimental source: thymus  
A;Accession: PC2246  
A;Molecule type: mRNA  
A;Residues: 1-62,'RFT' <KI2>  
A;Cross-references: DDBJ:D26113; NID:g468488; PIDN:BAA05109.1; PID:d1005651; PID:g468489  
A;Experimental source: liver  
C;Genetics:  
A;Introns: 62/1  
C;Superfamily: NGF receptor repeat homology  
C;Keywords: transmembrane protein  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-324/Product: Fas antigen #status predicted <MAT>  
F;44-79/Domain: NGF receptor repeat homology <NGF>  
F;81-124/Domain: NGF receptor repeat homology <NG4>  
F;171-188/Domain: transmembrane #status predicted <TMM>

Query Match 30.8%; Score 45; DB 2; Length 324;  
Best Local Similarity 26.4%; Pred. No. 66;  
Matches 14; Conservative 2; Mismatches 9; Indels 28; Gaps 2;

QY 3 EGLYQ-----GVPRAEPGTEA-----RRHYDEGVR 27  
||||| | | | | | :|:| : |  
Db 46 EGLYQVGFCCQPCQGERKVKDCTTSGGAPTCHEGEYTDKHYSDKCR 98

RESULT 37  
S38908  
glutathione reductase (NADPH) (EC 1.6.4.2) - common tobacco (fragment)  
C;Species: Nicotiana tabacum (common tobacco)  
C;Date: 22-Jan-1994 #sequence\_revision 10-Nov-1995 #text\_change 03-Nov-2000  
C;Accession: S38908  
R;Creissen, G.; Mullineaux, X.Y.Z.  
submitted to the EMBL Data Library, November 1993  
A;Reference number: S38908  
A;Accession: S38908  
A;Molecule type: mRNA  
A;Residues: 1-557 <CRE>  
A;Cross-references: EMBL:X76293; NID:g431954; PIDN:CAA53925.1; PID:g431955  
C;Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology  
C;Keywords: FAD; flavoprotein; NADP; oxidoreductase; redox-active disulfide  
F;83-528/Domain: dihydrolipoamide dehydrogenase homology <DLD>  
F;127-132/Disulfide bonds: redox-active #status predicted

Query Match 30.8%; Score 45; DB 2; Length 557;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 10; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 10 PRAEP--GTEARRHYD 23  
||||| | :|:|||||  
Db 64 PRAESSNGADAPRHYD 79

RESULT 38  
A65012  
hypothetical protein b2380 - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: A65012  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617  
A;Accession: A65012  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-565 <BLAT>  
A;Cross-references: GB:AE000326; GB:U00096; NID:g1788718; PIDN:AAC75439.1; PID:g17887  
A;Experimental source: strain K-12, substrain MGL655  
C;Superfamily: hypothetical protein b2380

Query Match 30.8%; Score 45; DB 1; Length 565;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 GEGLYQGVPRAPGTE 17  
||||| :|:|||||  
Db 532 GEGLH--IRRLEPGTE 545

RESULT 39  
F85880  
probable sensor protein Z3645 [imported] - Escherichia coli (strain O157:H7)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: F85880  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: F85880  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-565 <STO>  
A;Cross-references: GB:AE005174; NID:gl2516754; PIDN:AAG57506.1; GSPDB:GN00145; UWGP:  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z3645  
C;Superfamily: hypothetical protein b2380

Query Match 30.8%; Score 45; DB 2; Length 565;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 GEGLYQGVPRAPGTE 17  
||||| :|:|||||  
Db 532 GEGLH--IRRLEPGTE 545

RESULT 40  
A84206  
hypothetical protein porA [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: A84206  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483



A;Residues: 1-235 <KAW>  
A;Cross-references: DDBJ:AP0000058; NID:g5103388; PIDN:BAA79167.1; PID:d1042943; PID:g51042943  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE0254  
C;Superfamily: Aeropyrum pernix hypothetical protein APE0254

Query Match 30.5%; Score 44.5; DB 2; Length 235;  
Best Local Similarity 52.9%; Pred. No. 56;  
Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 8 GVPRAEPGTEARRHYDE 24  
I:II II III:I  
Db 182 GLRGEPAVR-RRHIEE 197

RESULT 45  
E70661  
probable PE protein - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: E70661  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: E70661  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-413 <COL>  
A;Cross-references: GB:Z83860; GB:AL123456; NID:g3261681; PIDN:CAB06157.1; PID:g1781247  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: PE  
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 30.5%; Score 44.5; DB 2; Length 413;  
Best Local Similarity 52.6%; Pred. No. 1e+02;  
Matches 10; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY 2 GEGLYQGVPRAPGTEARR 20  
I:II II III:I  
Db 328 GNAHQGNP-ADPGNSARR 345



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:12 ; Search time 37.02 Seconds  
(without alignments)  
14.692 Million cell updates/sec

Title: US-09-439-313-566  
Perfect score: 146  
Sequence: 1 VGEGLYQGVPRAPGTEARRHYDEGVR 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : Issued\_Patents.AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131	89.7	27	4	US-09-071-710-39
2	131	89.7	255	4	US-09-071-710-36
3	46.5	31.8	140	1	US-08-664-596B-2
4	46.5	31.8	140	2	US-08-738-367-2
5	46.5	31.8	240	2	US-08-738-367-7
6	46	31.5	529	1	US-08-172-331B-14
7	46	31.5	529	1	US-08-706-037-25
8	46	31.5	529	2	US-09-005-397-25
9	46	31.5	529	2	US-09-032-315-5
10	46	31.5	529	2	US-08-993-318A-5
11	46	31.5	529	4	US-09-399-886-5
12	46	31.5	529	4	US-09-396-260-5
13	46	31.5	564	2	US-08-865-337A-1
14	46	31.5	610	2	US-08-865-337A-6
15	43.5	29.8	4544	1	US-08-469-486-52
16	43.5	29.8	4544	2	US-08-469-658-52
17	43	29.5	113	2	US-08-563-892A-30
18	43	29.5	323	2	US-08-435-149-2
19	43	29.5	324	1	US-08-310-416A-14
20	43	29.5	324	2	US-08-888-171-14
21	43	29.5	577	2	US-08-435-149-3
22	42.5	29.1	445	1	US-08-674-168-30
23	42.5	29.1	445	3	US-08-985-908-20
24	42.5	29.1	445	3	US-08-852-730-5
25	42.5	29.1	445	4	US-08-985-916-12
26	42	28.8	348	3	US-08-652-265-4
27	42	28.8	348	3	US-08-652-265-8

28	42	28.8	348	4	US-08-834-497A-4	Sequence 4, Appli
29	42	28.8	348	4	US-08-834-497A-8	Sequence 8, Appli
30	42	28.8	348	4	US-09-503-444A-4	Sequence 4, Appli
31	42	28.8	348	4	US-09-503-444A-8	Sequence 8, Appli
32	42	28.8	350	2	US-08-960-022-8	Sequence 8, Appli
33	42	28.8	823	3	US-08-481-435-4	Sequence 4, Appli
34	42	28.8	836	3	US-08-481-435-9	Sequence 9, Appli
35	42	28.8	844	3	US-08-481-435-7	Sequence 7, Appli
36	42	28.8	844	3	US-08-481-435-8	Sequence 8, Appli
37	41	28.1	30	4	US-08-777-708C-2	Sequence 2, Appli
38	41	28.1	60	4	US-08-777-708C-23	Sequence 23, Appli
39	41	28.1	250	1	US-08-562-311-2	Sequence 2, Appli
40	41	28.1	250	3	US-08-946-914-10	Sequence 10, Appli
41	41	28.1	290	4	US-09-286-691-29	Sequence 29, Appli
42	41	28.1	310	1	US-08-078-683A-3	Sequence 3, Appli
43	41	28.1	310	2	US-08-488-199-4	Sequence 4, Appli
44	41	28.1	499	2	US-09-032-315-3	Sequence 3, Appli
45	41	28.1	499	2	US-08-993-318A-3	Sequence 3, Appli
46	41	28.1	499	4	US-09-399-886-3	Sequence 3, Appli
47	41	28.1	499	4	US-09-396-260-3	Sequence 3, Appli
48	41	28.1	502	4	US-09-413-814-106	Sequence 106, App
49	41	28.1	518	3	US-08-999-723-2	Sequence 2, Appli
50	41	28.1	518	4	US-09-434-427-2	Sequence 2, Appli
51	41	28.1	519	1	US-08-462-484-4	Sequence 4, Appli
52	41	28.1	519	1	US-08-441-147-4	Sequence 4, Appli
53	41	28.1	519	5	PCT-US95-07536-4	Sequence 4, Appli
54	41	28.1	1525	4	US-09-396-651B-1	Sequence 1, Appli
55	40.5	27.7	327	4	US-09-290-640-66	Sequence 66, Appli
56	40	27.4	493	1	US-07-615-448A-7	Sequence 7, Appli
57	40	27.4	493	1	US-08-196-361-7	Sequence 7, Appli
58	40	27.4	493	2	US-08-446-934-7	Sequence 7, Appli
59	40	27.4	493	2	US-08-448-128-7	Sequence 7, Appli
60	40	27.4	493	3	US-08-948-703-7	Sequence 7, Appli
61	40	27.4	769	4	US-09-320-878-12	Sequence 12, Appli
62	40	27.4	776	2	US-08-700-607-5	Sequence 5, Appli
63	40	27.4	1244	4	US-08-938-291A-5	Sequence 5, Appli
64	39.5	27.1	438	5	PCT-US92-09325-4	Sequence 4, Appli
65	39.5	27.1	445	1	US-08-700-359-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-071-710-39  
; Sequence 39, Application US/09071710  
; Patent No. 6130043  
; GENERAL INFORMATION:  
; APPLICANT: BILLING-MEDEL, PATRICIA  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GORDON, JULIAN  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: HODGES, STEVEN C.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: KRATOCHVIL, JON D.  
; APPLICANT: ROBERTS-RAPP, LISA  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STROUPE, STEPHEN D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE PROSTATE  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/071,710
; APPLICATION NUMBER: US/09/071,710
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/850,713
; APPLICATION NUMBER: 08/850,713
; FILING DATE: 02-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6130043e
; US-09-071-710-39

```

```

Query Match      89.7%; Score 131; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 GLYQGVPRAPGTEARRHYDEGVR 27
Db 1 GLYQGVPRAPGTEARRHYDEGVR 24

```

```

RESULT 2
US-09-071-710-36
; Sequence 36, Application US/09071710
; Patent No. 6130043
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; FOR DETECTING DISEASES OF THE PROSTATE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,710
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/850,713
; FILING DATE: 02-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 255 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6130043e
; US-09-071-710-36

```

```

Query Match      89.7%; Score 131; DB 4; Length 255;
Best Local Similarity 100.0%; Pred. No. 5.4e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 GLYQGVPRAPGTEARRHYDEGVR 27
Db 1 GLYQGVPRAPGTEARRHYDEGVR 24

```

```

RESULT 3
US-08-664-596B-2
; Sequence 2, Application US/08664596B
; Patent No. 5807703
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: MCCOY, John
; APPLICANT: Lavalie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; APPLICANT: Bowman, Michael
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/664,596B
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; STRANDEDNESS:

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-564-596B-2

Query Match 31.8%; Score 46.5; DB 1; Length 140;
Best Local Similarity 47.4%; Pred. No. 8.4;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYQGVPRAPGTEARRHYD 23
:: ||| |||: | | :|
Db 83 VWSGVPAEMPGTK-RHHFD 100

RESULT 4
US-08-738-367-2
; Sequence 2, Application US/08738367
; Patent No. 5827688
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: LaVallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,367
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-738-367-2

Query Match 31.8%; Score 46.5; DB 2; Length 140;
Best Local Similarity 47.4%; Pred. No. 8.4;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYQGVPRAPGTEARRHYD 23
:: ||| |||: | | :|
Db 83 VWSGVPAEMPGTK-RHHFD 100

RESULT 5
US-08-738-367-7
; Sequence 7, Application US/08738367
```

```

; Patent No. 5827688
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: LaVallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,367
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-738-367-7

Query Match 31.8%; Score 46.5; DB 2; Length 240;
Best Local Similarity 47.4%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYQGVPRAPGTEARRHYD 23
:: ||| |||: | | :|
Db 83 VWSGVPAEMPGTK-RHHFD 100

RESULT 6
US-08-172-331B-14
; Sequence 14, Application US/08172331B
; Patent No. 5480801
; GENERAL INFORMATION:
; APPLICANT: Wahleithner, Jill A.
; APPLICANT: Christensen, Bjoern E.
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5480801o No. 5480801disk of No. 5480801th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/172,331B  
FILING DATE: 22-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: US 08/122,230  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA: US 08/122,827  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA: US 08/162,827  
FILING DATE: 03-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lowney Dr., Karen A.  
REGISTRATION NUMBER: 31,274  
REFERENCE/DOCKET NUMBER: 4052.020-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-172-331B-14

Query Match 31.5%; Score 46; DB 1; Length 529;  
Best Local Similarity 66.7%; Pred. NO. 46;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17  
Db 314 YEGAPNAEPTTE 325

RESULT 7  
US-08-706-037-25  
Sequence 25, Application US/08706037  
Patent No. 5770419  
GENERAL INFORMATION:  
APPLICANT: Xu, Feng  
APPLICANT: Berk, Randy M.  
APPLICANT: Wahleithner, Jill A.  
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH  
ENHANCED ACTIVITY  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 57704190 No. 5770419disk of No. 5770419th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/706,037  
FILING DATE: 30-AUG-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,800  
FILING DATE: 1-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.

REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4526.200-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-706-037-25

Query Match 31.5%; Score 46; DB 1; Length 529;  
Best Local Similarity 66.7%; Pred. NO. 46;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17  
Db 314 YEGAPNAEPTTE 325

RESULT 8  
US-09-005-397-25  
Sequence 25, Application US/09005397  
Patent No. 5972670  
GENERAL INFORMATION:  
APPLICANT: Xu, Feng  
APPLICANT: Berk, Randy M.  
APPLICANT: Wahleithner, Jill A.  
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH  
ENHANCED ACTIVITY  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 59726700 No. 5972670disk of No. 5972670th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005,397  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/706,037  
FILING DATE: 30-AUG-1996  
APPLICATION NUMBER: US 60/002,800  
FILING DATE: 1-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 4526.200-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-005-397-25

Query Match 31.5%; Score 46; DB 2; Length 529;  
Best Local Similarity 66.7%; Pred. NO. 46;



Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17  
|:| | ||| ||

Db 314 YEGAPNAEPTTE 325

RESULT 9

US-09-032-315-5

; Sequence 5, Application US/09032315

; Patent No. 5985818

; GENERAL INFORMATION:

; APPLICANT: Svendsen, Allan

; APPLICANT: Xu, Feng

; TITLE OF INVENTION: LACCASE MUTANTS

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5985818o No. 5985818disk of No. 5985818th America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/032,315

; FILING DATE: 27-FEB-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Rozek, Carol

; REGISTRATION NUMBER: 36,993

; REFERENCE/DOCKET NUMBER: 5200,200-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 529 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-032-315-5

Query Match 31.5%; Score 46; DB 2; Length 529;

Best Local Similarity 66.7%; Pred. No. 46;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17  
|:| | ||| ||

Db 314 YEGAPNAEPTTE 325

RESULT 10

US-08-993-318A-5

; Sequence 5, Application US/08993318A

; Patent No. 5998353

; GENERAL INFORMATION:

; APPLICANT: Pedersen, Anders

; APPLICANT: Svendsen, Allan

; APPLICANT: Schneider, Palle

; APPLICANT: Rasmussen, Grethe

; APPLICANT: Cherry, Joel

; TITLE OF INVENTION: LACCASE MUTANTS

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5998353o No. 5998353disk of No. 5998353th America

; STREET: 405 Lexington Avenue

; CITY: New York

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/993,318A

; FILING DATE: December 18, 1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta A.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 5032,200-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 529 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-993-318A-5

Query Match 31.5%; Score 46; DB 2; Length 529;

Best Local Similarity 66.7%; Pred. No. 46;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17  
|:| | ||| ||

Db 314 YEGAPNAEPTTE 325

RESULT 11

US-09-399-886-5

; Sequence 5, Application US/09399886

; Patent No. 6140092

; GENERAL INFORMATION:

; APPLICANT: Pedersen, Anders

; APPLICANT: Svendsen, Allan

; APPLICANT: Schneider, Palle

; APPLICANT: Rasmussen, Grethe

; APPLICANT: Cherry, Joel

; TITLE OF INVENTION: LACCASE MUTANTS

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6140092o No. 6140092disk of No. 6140092th America

; STREET: 405 Lexington Avenue

; CITY: New York

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/399,886

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/993,318

; FILING DATE: December 18, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta A.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 5032,200-US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-399-886-5

Query Match 31.5%; Score 46; DB 4; Length 529;  
Best Local Similarity 66.7%; Pred. No. 46;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17  
DB 314 YEGAPNAEPTTE 325

RESULT 12  
US-09-396-260-5  
Sequence 5, Application US/09396260  
Patent No. 6184015  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 61840150 No. 6184015disk of No. 6184015th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/396,260  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,315  
FILING DATE: 27-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 5200.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-396-260-5

Query Match 31.5%; Score 46; DB 4; Length 529;  
Best Local Similarity 66.7%; Pred. No. 46;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17

DB 314 YEGAPNAEPTTE 325

RESULT 13  
US-08-865-337A-1  
Sequence 1, Application US/08865337A  
Patent No. 5972649  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Covitz, Peter  
APPLICANT: Tang, Y. Tom  
TITLE OF INVENTION: HUMAN MULTIPLE ENDOCRINE NEOPLASIA TYPE  
NUMBER OF SEQUENCES: 1 PROTEIN  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/865,337A  
FILING DATE: Herewith  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0305 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 564 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAINOT14  
CLONE: Consensus  
US-08-865-337A-1

Query Match 31.5%; Score 46; DB 2; Length 564;  
Best Local Similarity 52.9%; Pred. No. 49;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEA 18  
DB 461 GQGA VSGPPRKPPTVA 477

RESULT 14  
US-08-865-337A-6  
Sequence 6, Application US/08865337A  
Patent No. 5972649  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Covitz, Peter  
APPLICANT: Tang, Y. Tom  
APPLICANT: Murry, Lynn E.

;; TITLE OF INVENTION: HUMAN MULTIPLE ENDOCRINE NEOPLASIA TYPE  
;; TITLE OF INVENTION: 1 PROTEIN  
;; NUMBER OF SEQUENCES: 6  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
;; STREET: 3174 Porter Drive  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94304  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/865,337A  
;; FILING DATE: Herewith  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Billings, Lucy J.  
;; REGISTRATION NUMBER: 36,749  
;; REFERENCE/DOCKET NUMBER: PF-0305 US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-855-0555  
;; TELEFAX: 415-845-4166  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 610 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; LIBRARY: GenBank  
;; CLONE: 1945387  
;; US-08-865-337A-6

Query Match 31.5%; Score 46; DB 2; Length 610;  
Best Local Similarity 52.9%; Pred. No. 54;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEA 18  
|:| | | | | | | | | |  
Db 507 GQGAVGPPRKPPTGTA 523

RESULT 15  
US-08-469-486-52  
; Sequence 52, Application US/08469486  
; Patent No. 5739281  
; GENERAL INFORMATION:  
; APPLICANT: Thøgersen, Hans Christian  
; APPLICANT: Holtet, Thor Las  
; APPLICANT: Etzerodt, Michael  
; TITLE OF INVENTION: Improved method for the refolding of  
; TITLE OF INVENTION: proteins  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version  
;; SOFTWARE: #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/469,486  
;; FILING DATE:  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/192,060  
;; FILING DATE: February 4, 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Paul T. Clark  
;; REGISTRATION NUMBER: 30,162  
;; REFERENCE/DOCKET NUMBER: 06363/002001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617 542 5070  
;; TELEFAX: 617 542 8906  
;; TELEX: 200154  
;; INFORMATION FOR SEQ ID NO: 52:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4544 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-469-486-52

Query Match 29.8%; Score 43.5; DB 1; Length 4544;  
Best Local Similarity 41.7%; Pred. No. 1.3e+03;  
Matches 10; Conservative 3; Mismatches 8; Indels 3; Gaps 1;

QY 6 YQGVPRAPGTEARRH---YDEGV 26  
|:| | | | | | | | | |  
Db 3926 YRSLPPAAPPTTSNRHRRQIDRGV 3949

RESULT 16  
US-08-469-658-52  
; Sequence 52, Application US/08469658  
; Patent No. 5917018  
; GENERAL INFORMATION:  
; APPLICANT: Thøgersen, Hans Christian  
; APPLICANT: Holtet, Thor Las  
; APPLICANT: Etzerodt, Michael  
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF  
; TITLE OF INVENTION: PROTEINS  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version  
; SOFTWARE: #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,658  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/192,060  
; FILING DATE: February 4, 1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul T. Clark  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 06363/002002  
; TELECOMMUNICATION INFORMATION:

QY 3 EGLYQGVPRAEPTARRH 21  
:||||: | || |  
Pb 88 DGLYQGLSTATKPTDALH 106

US-08-435-149-2

COMPLANCE Fish & Wildlife Richardson P.C.  
ADDRESS 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB





```

; Patent No. 5804414
; GENERAL INFORMATION:
; APPLICANT: MORIYA, Mika
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: YOKOZEKI, Kenzo
; APPLICANT: HIRANO, Seiko
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: IZUI, Masako
; APPLICANT: SUGIMOTO, Masakazu
; TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
; TITLE OF INVENTION: ARTIFICIAL TRANSPOSON
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C.
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,168
; FILING DATE: 01-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-166541
; FILING DATE: 30-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-810-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELETEXT: 24855 OPAT UR
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-674-168-30

```

```

Query Match 29.1%; Score 42.5; DB 1; Length 445;
Best Local Similarity 52.4%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 2 GEGLYQGV-PRAEPGTEARRH 21
Db 154 GEGKIQDVLRVKGPIEAHTH 174

RESULT 23
US-08-985-908-20
; Sequence 20, Application US/08985908
; Patent No. 6004773
; GENERAL INFORMATION:
; APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI NA
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; COUNTRY: USA
; ZIP: 22202

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,908
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-325659
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-985-908-20

```

```

Query Match 29.1%; Score 42.5; DB 3; Length 445;
Best Local Similarity 52.4%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 2 GEGLYQGV-PRAEPGTEARRH 21
Db 154 GEGKIQDVLRVKGPIEAHTH 174

RESULT 24
US-08-852-730-5
; Sequence 5, Application US/08852730
; Patent No. 6090597
; GENERAL INFORMATION:
; APPLICANT: SEIKO HIRANO, MASAKAZU SUGIMOTO, EIICHI NAKANO,
; APPLICANT: MASAKO IZUI, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI
; APPLICANT: NAKAMATSU
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER AND NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; ZIP: 22026
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,730
; FILING DATE: 05-07-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-142812
; FILING DATE: 05-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids

```

;  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-852-730-5

Query Match 29.1%; Score 42.5; DB 3; Length 445;  
Best Local Similarity 52.4%; Pred. No. 1.2e+02;  
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 2 GEGLYQGV-PRAEPGTEARRH 21  
||| | | | :|| | |  
Db 154 GEGKIQDVLRVKPGIEATH 174

RESULT 25  
US-08-985-916-12  
; Sequence 12, Application US/08985916  
; Patent No. 6221636  
; GENERAL INFORMATION:  
; APPLICANT: ATSUSHI HAYAKAWA, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI  
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
; CITY: ARLINGTON  
; COUNTRY: VA  
; ZIP: 22152

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/985,916  
; FILING DATE: 05-DEC-1997  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-325658  
; FILING DATE: 05-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NORMAN F. OBLON  
; REGISTRATION NUMBER: 24,618  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 445 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-985-916-12

Query Match 29.1%; Score 42.5; DB 4; Length 445;  
Best Local Similarity 52.4%; Pred. No. 1.2e+02;  
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 2 GEGLYQGV-PRAEPGTEARRH 21  
||| | | | :|| | |  
Db 154 GEGKIQDVLRVKPGIEATH 174

RESULT 26  
US-08-652-265-4  
; Sequence 4, Application US/08652265  
; Patent No. 6025130  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Winston J.  
; APPLICANT: Drayna, Dennis T.  
; APPLICANT: Feder, John N.

; APPLICANT: Gnirke, Andreas  
; APPLICANT: Ruddy, David  
; APPLICANT: Tsuchihashi, Zenta  
; APPLICANT: Wolff, Roger K.  
; TITLE OF INVENTION: Hereditary Hemochromatosis Gene  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,265  
; FILING DATE: 23-MAY-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M.  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 17957-000500  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 348 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-652-265-4

Query Match 28.8%; Score 42; DB 3; Length 348;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 5; Indels 8; Gaps 2;

QY 2 GEGLYQG-----VPRAEPGTEARRHY 22  
|:| ||| || ||| |  
Db 260 GDGTYQGWITLAVP---PGEEQRYTY 282

RESULT 27  
US-08-652-265-8  
; Sequence 8, Application US/08652265  
; Patent No. 6025130  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Winston J.  
; APPLICANT: Drayna, Dennis T.  
; APPLICANT: Feder, John N.  
; APPLICANT: Gnirke, Andreas  
; APPLICANT: Ruddy, David  
; APPLICANT: Tsuchihashi, Zenta  
; APPLICANT: Wolff, Roger K.  
; TITLE OF INVENTION: Hereditary Hemochromatosis Gene  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

Query Match: 28.8%; Score 42; DB 4; Length 348;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 5; Indels

Query Match: 28.8%; Score 42; DB 4; Length 348;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 5; Indels

QY 2 GEGLYQG----VPRAEPGTEARRHY 22  
| : | | | | | | | | | |  
pb 260 GDGTyOGWITLAVP---PGEEQRyTY 282

RESULT 29  
 US-08-834-497A-8  
 ; Sequence 8, Application US/08834497A  
 ; Patent No. 6140305  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Thomas, Winston J.  
 ; APPLICANT: Drayna, Dennis T.  
 ; APPLICANT: Feder, John N.  
 ; APPLICANT: Gnirke, Andreas  
 ; APPLICANT: Ruddy, David  
 ; APPLICANT: Tsuchihashi, Zenta  
 ; APPLICANT: Wolff, Roger K.  
 ; TITLE OF INVENTION: HEREDITARY HEMOCHROMATOSIS GENE PRODUCTS  
 ; NUMBER OF SEQUENCES: 76  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: pennie & Edmonds LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2811  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: Windows 95  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0b  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/834,497A  
 ; FILING DATE: 04-APR-1997  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/652,265  
 ; FILING DATE: 23-MAY-1996  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/632,673  
 ; FILING DATE: 16-APR-1996  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/630,912  
 ; FILING DATE: 04-APR-1996  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Poissant, Brian M.  
 ; REGISTRATION NUMBER: 28,462  
 ; REFERENCE/DOCKET NUMBER: 8907-0056-999  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-493-4935  
 ; TELEFAX: 650-493-5556



```

;
; TELEX: 66141 PENNIE
;
; INFORMATION FOR SEQ ID NO: 8:
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 348 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; US-08-834-497A-8

```

Query Match 28.8%; Score 42; DB 4; Length 348;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 5; Indels

Qy      2    GEGLYQG-----VPRAEPGTEARRHY    22  
         | : | | |       | |     | | | |  
Db    260   GDGTYYQGWTLPV---PGEEQRITY    282

RESULT 30  
US-09-503-444A-4  
; Sequence 4, Application US/09503444A  
; Patent No. 6228594  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Winston J.  
; APPLICANT: Drayna, Dennis T.  
; APPLICANT: Feder, John N.  
; APPLICANT: Gnirke, Andreas  
; APPLICANT: Ruddy, David  
; APPLICANT: Tsuchihashi, Zenta  
; APPLICANT: Wolff, Roger K.  
; TITLE OF INVENTION: Hereditary Hemochromatosis Gene  
; NUMBER OF SEQUENCES: 44

Query Match 28.8%; Score 42; DB 4; Length 348;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 5; Indels

QY	2	GEGLYQG----	VPRAEPGTEARRHY	22
		:		
Db	260	GDGTYQGWITLAVP---	PGEEQRYTY	282

RESULT 31  
US-09-503-444A-8  
; Sequence 8, Application US/09503444A  
; Patent No. 6228594  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Winston J.  
; APPLICANT: Drayna, Dennis T.  
; APPLICANT: Feder, John N.  
; APPLICANT: Gnirke, Andreas  
; APPLICANT: Ruddy, David  
; APPLICANT: Tsuchihashi, Zenta  
; APPLICANT: Wolff, Roger K.  
; TITLE OF INVENTION: Hereditary Hemochromatosis Gene  
; NUMBER OF SEQUENCES: 44

Query Match 28.8%; Score 42; DB 4; Length 348;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 12: Conservative 1; Mismatches 5; Indels

QY 2 GEGLYQG-----VPRAEPGTEARRHY 22  
|:| ||| || ||| | | | |

Db 260 GDGTQGWITLAVP---PGEEQRYTY 282

RESULT 32

US-08-960-022-8  
; Sequence 8, Application US/08960022  
; Patent No. 5976837  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LaVallie, Edward R.  
; APPLICANT: Racie, Lisa A.  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Agostino, Michael J.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; TITLE OF INVENTION: ENCODING THEM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/960,022  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sprunger, Suzanne A.  
; REGISTRATION NUMBER: 41,323  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8284  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 350 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-960-022-8

Query Match 28.8%; Score 42; DB 2; Length 350;  
Best Local Similarity 47.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 EGLYQGVPRAPGTEAR 19  
Db 41 ENLVRSPSGEPGREK 57

RESULT 33

US-08-481-435-4  
; Sequence 4, Application US/08481435  
; Patent No. 6027906  
; GENERAL INFORMATION:  
; APPLICANT: Balganes, Tanjore S  
; APPLICANT: Town, Christine  
; TITLE OF INVENTION: No. 6027906el Polypeptides  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: White & Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York

STATE: New York  
COUNTRY: United States  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,435  
FILING DATE: 10-JUL-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IN 580/MAS/94  
FILING DATE: 01-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9404072-2  
FILING DATE: 24-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Steiner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-151  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 823 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-481-435-4

Query Match 28.8%; Score 42; DB 3; Length 823;  
Best Local Similarity 42.9%; Pred. No. 3e+02;  
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEARRHY 22  
Db 627 GKVLQSFQPAERAVPAQAAY 647

RESULT 34

US-08-481-435-9  
; Sequence 9, Application US/08481435  
; Patent No. 6027906  
; GENERAL INFORMATION:  
; APPLICANT: Balganes, Tanjore S  
; APPLICANT: Town, Christine  
; TITLE OF INVENTION: No. 6027906el Polypeptides  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: White & Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,435  
FILING DATE: 10-JUL-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IN 580/MAS/94  
FILING DATE: 01-JUL-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9404072-2

```

; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 836 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; IMMEDIATE SOURCE:
; CLONE: PARC0469 PBP 1B del 8
;
US-08-481-435-9

Query Match 28.8%; Score 42; DB 3; Length 836;
Best Local Similarity 42.9%; Pred. No. 3e+02;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAEPTGTEARRHY 22
   I: ||| I: || I: |
Db 640 GKVLYQSFPQAERAVPAQAAY 660

RESULT 35
US-08-481-435-7
; Sequence 7, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,435
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IN 580/MAS/94
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9404072-2
; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 844 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; IMMEDIATE SOURCE:
; CLONE: PARC0468 PBP 1B QQLL
;
US-08-481-435-8

Query Match 28.8%; Score 42; DB 3; Length 844;
Best Local Similarity 42.9%; Pred. No. 3.1e+02;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAEPTGTEARRHY 22
   I: ||| I: || I: |
Db 648 GKVLYQSFPQAERAVPAQAAY 668

RESULT 36
US-08-481-435-8
; Sequence 8, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,435
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IN 580/MAS/94
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9404072-2
; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 844 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; IMMEDIATE SOURCE:
; CLONE: PARC0468 PBP 1B QQLL
;
US-08-481-435-8

Query Match 28.8%; Score 42; DB 3; Length 844;
Best Local Similarity 42.9%; Pred. No. 3.1e+02;

```

Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAEPTGTEARRHY 22  
|: ||| |::| |:  
Db 648 GKVLQSFQAEARAVPAQRAY 668

RESULT 37  
US-08-777-708C-2  
; Sequence 2, Application US/08777708C  
; Patent No. 6159934  
; GENERAL INFORMATION:  
; APPLICANT: Pescovitz, Ora H.  
; TITLE OF INVENTION: USE OF GHRH-RP TO STIMULATE STEM CELL FACTOR PRODUCTION  
; FILE REFERENCE: 7037-175/1U-62  
; CURRENT APPLICATION NUMBER: US/08/777,708C  
; CURRENT FILING DATE: 1996-12-20  
; PRIOR APPLICATION NUMBER: US 60/008,933  
; PRIOR FILING DATE: 1995-12-20  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 30  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-08-777-708C-2

Query Match 28.1%; Score 41; DB 4; Length 30;  
Best Local Similarity 43.8%; Pred. No. 9.4;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EGLYQGVPRAEPTGTEA 18  
|: ||| |::| |:  
Db 15 ESILQGFPRMKPSADA 30

RESULT 38  
US-08-777-708C-23  
; Sequence 23, Application US/08777708C  
; Patent No. 6159934  
; GENERAL INFORMATION:  
; APPLICANT: Pescovitz, Ora H.  
; TITLE OF INVENTION: USE OF GHRH-RP TO STIMULATE STEM CELL FACTOR PRODUCTION  
; FILE REFERENCE: 7037-175/1U-62  
; CURRENT APPLICATION NUMBER: US/08/777,708C  
; CURRENT FILING DATE: 1996-12-20  
; PRIOR APPLICATION NUMBER: US 60/008,933  
; PRIOR FILING DATE: 1995-12-20  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 60  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-08-777-708C-23

Query Match 28.1%; Score 41; DB 4; Length 60;  
Best Local Similarity 43.8%; Pred. No. 21;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 EGLYQGVPRAEPTGTEA 18  
|: ||| |::| |:  
Db 45 ESILQGFPRMKPSADA 60

RESULT 39  
US-08-562-311-2  
; Sequence 2, Application US/08562311  
; Patent No. 5801002  
; GENERAL INFORMATION:  
; APPLICANT: RAZ, AVRAHAM

; TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY  
; OF METASTASIS IN A CELL SAMPLE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dykema Gossett  
; STREET: STE 505 N. Woodward  
; CITY: Bloomfield Hills  
; STATE: MI  
; COUNTRY: U.S.  
; ZIP: 48304  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/562,311  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,225  
; FILING DATE:  
; APPLICATION NUMBER: US 07/681,242  
; FILING DATE: 04-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/294,249  
; FILING DATE: 01-JUN-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KELLY, ROBERT L.  
; REGISTRATION NUMBER: 31,843  
; REFERENCE/DOCKET NUMBER: 61,686-  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 810-540-0849  
; TELEFAX: 810-540-0763  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 250 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-562-311-2

Query Match 28.1%; Score 41; DB 1; Length 250;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 GLYQGVPRAEPTGTEA 18  
|: ||| |::| |:  
Db 61 GAYHGAPGAYPGAPA 75

RESULT 40  
US-08-946-914-10  
; Sequence 10, Application US/08946914  
; Patent No. 6027916  
; GENERAL INFORMATION:  
; APPLICANT: Ni, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS



; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/946,914  
; FILING DATE: Herewith  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/028,093  
; FILING DATE: 09-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 250 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-946-914-10

Query Match 28.1%; Score 41; DB 3; Length 250;  
Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
Matches 8; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 GLYQGVPRAPGTEA 18  
| | | | | | | |  
Db 61 GAYHGAPGAYPGAPA 75

## RESULT 41

US-09-286-691-29  
; Sequence 29, Application US/09286691  
; Patent No. 6190189  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xin-Liang  
; APPLICANT: Ljungdahl, Lars G.  
; APPLICANT: Chen, Huizhong  
; TITLE OF INVENTION: Cellulases and Coding Sequences  
; FILE REFERENCE: 42-96  
; CURRENT APPLICATION NUMBER: US/09/286,691  
; CURRENT FILING DATE: 1999-04-05  
; EARLIER APPLICATION NUMBER: US 60/027,883  
; EARLIER FILING DATE: 1996-10-04  
; EARLIER APPLICATION NUMBER: PCT US97/18008  
; EARLIER FILING DATE: 1997-10-03  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: Streptomyces Ksm-9  
US-09-286-691-29

Query Match 28.1%; Score 41; DB 4; Length 290;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 4 GLYQGVPRAPGTEARRHYDEG 25  
| | : | | | | | | | |  
Db 145 GLAEAVTDANP--EARVYIDVG 164

## RESULT 42

US-08-078-683A-3  
; Sequence 3, Application US/08078683A  
; Patent No. 5486599  
; GENERAL INFORMATION:

; APPLICANT: Saunders, Scott  
; APPLICANT: Bernfield, Merton  
; APPLICANT: Kato, Masato  
; TITLE OF INVENTION: Construction and Use of Synthetic  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII (text)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/078,683A  
; FILING DATE: 17-JUN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vincent, Matthew P.  
; REGISTRATION NUMBER: 36,709  
; REFERENCE/DOCKET NUMBER: CME-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 310 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-08-078-683A-3

Query Match 28.1%; Score 41; DB 1; Length 310;  
Best Local Similarity 45.5%; Pred. No. 1.4e+02;  
Matches 10; Conservative 2; Mismatches 6; Indels 4; Gaps 1;

QY 2 GEGLYQG---VPRAEPGTEAR 19  
| | | : | : | | | | |  
Db 96 GEGPKEGEAVVLPEVEPLTAR 117

## RESULT 43

US-08-488-199-4  
; Sequence 4, Application US/08488199  
; Patent No. 5851993  
; GENERAL INFORMATION:  
; APPLICANT: Jalkanen, Markku  
; APPLICANT: Mali, Markku  
; TITLE OF INVENTION: Suppression of Tumor Cell Growth By  
; TITLE OF INVENTION: Syndecan-1 Ectodomain  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX  
; STREET: 1100 New York Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,199  
; FILING DATE: 07-JUN-1995

CLASSIFICATION: 514  
PRIOR APPLICATION DATA: US 08/258,862  
APPLICATION NUMBER: 13-JUN-1994  
FILING DATE: 13-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 1102.0130001  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-488-199-4

Query Match 28.1%; Score 41; DB 2; Length 310;  
Best Local Similarity 45.5%; Pred. NO. 1.4e+02;  
Matches 10; Conservative 2; Mismatches 6; Indels 4; Gaps 1;

QY 2 GEGLYQG----VPRAEPGTAR 19  
Db 96 GEGPKEGEAVVLPVEPGLTAR 117

RESULT 44  
US-09-032-315-3  
Sequence 3, Application US/09032315  
Patent No. 5985818  
GENERAL INFORMATION:  
APPLICANT: Svendsen, Allan  
APPLICANT: Xu, Feng  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5985818 No. 5985818disk of No. 5985818th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,315  
FILING DATE: 27-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 5200.200-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 499 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-032-315-3

Query Match 28.1%; Score 41; DB 2; Length 499;  
Best Local Similarity 72.7%; Pred. NO. 2.4e+02;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAEPT 16  
Db 281 YQGAPVAEPTT 291

RESULT 45  
US-08-993-318A-3  
Sequence 3, Application US/08993318A  
Patent No. 5998353  
GENERAL INFORMATION:  
APPLICANT: Pedersen, Anders  
APPLICANT: Svendsen, Allan  
APPLICANT: Schneider, Palle  
APPLICANT: Rasmussen, Grethe  
APPLICANT: Cherry, Joel  
TITLE OF INVENTION: LACCASE MUTANTS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5998353 No. 5998353disk of No. 5998353th America  
STREET: 405 Lexington Avenue  
CITY: New York  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993,318A  
FILING DATE: December 18, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta A.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 5032.200-US  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 499 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-993-318A-3

Query Match 28.1%; Score 41; DB 2; Length 499;  
Best Local Similarity 72.7%; Pred. NO. 2.4e+02;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAEPT 16  
Db 281 YQGAPVAEPTT 291

Search completed: June 28, 2001, 11:56:14  
Job time: 195 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:14 ; Search time 37.02 Seconds  
(without alignments)  
5.442 Million cell updates/sec

Title: US-09-439-313-573  
Perfect score: 45  
Sequence: 1 MGSIVQLSQS 10

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	255	4	US-09-071-710-36
2	31	68.9	468	1	US-08-459-287-2
3	31	68.9	468	1	US-08-459-287-3
4	31	68.9	468	4	US-08-764-870-8
5	29	64.4	240	2	US-08-800-198-8
6	29	64.4	240	4	US-09-296-595-8
7	29	64.4	241	1	US-08-235-838-5
8	29	64.4	241	1	US-08-235-838-11
9	29	64.4	241	2	US-08-465-473B-5
10	29	64.4	241	2	US-08-887-352B-11
11	29	64.4	248	2	US-08-887-352B-22
12	29	64.4	248	2	US-08-887-352B-23
13	29	64.4	248	4	US-09-109-207C-22
14	29	64.4	248	4	US-09-109-207C-23
15	29	64.4	249	2	US-08-797-689-18
16	29	64.4	271	2	US-08-400-115-4
17	29	64.4	301	2	US-08-661-052-14
18	29	64.4	301	2	US-08-656-906-25
19	29	64.4	301	4	US-09-217-847-25
20	29	64.4	355	3	US-08-875-811-41
21	29	64.4	355	3	US-08-875-811-49
22	29	64.4	355	3	US-08-875-811-64
23	29	64.4	358	3	US-08-875-811-45
24	29	64.4	358	3	US-08-875-811-51
25	29	64.4	360	3	US-08-875-811-47
26	29	64.4	379	3	US-08-875-811-43
27	29	64.4	553	2	US-08-661-052-16

28	29	64.4	637	1	US-08-235-838-14	Sequence 14, Appl
29	29	64.4	637	1	US-08-235-838-16	Sequence 16, Appl
30	29	64.4	637	2	US-08-465-473B-14	Sequence 14, Appl
31	29	64.4	637	2	US-08-465-473B-16	Sequence 16, Appl
32	29	64.4	711	1	US-08-235-838-7	Sequence 7, Appli
33	29	64.4	711	2	US-08-465-473B-7	Sequence 7, Appli
34	29	64.4	1070	2	US-08-633-770A-2	Sequence 2, Appli
35	28	62.2	241	4	US-08-902-486-13	Sequence 13, Appl
36	28	62.2	321	4	US-08-748-506-10	Sequence 10, Appl
37	28	62.2	321	4	US-08-748-506-18	Sequence 18, Appl
38	28	62.2	496	4	US-08-902-486-15	Sequence 15, Appl
39	28	62.2	612	2	US-08-933-750C-36	Sequence 36, Appl
40	28	62.2	612	4	US-09-234-613-36	Sequence 36, Appl
41	28	62.2	680	1	US-07-674-287B-2	Sequence 2, Appli
42	28	62.2	680	2	US-08-436-900A-2	Sequence 2, Appli
43	28	62.2	682	2	US-08-436-900A-4	Sequence 4, Appli
44	28	62.2	910	4	US-08-460-269C-2	Sequence 2, Appli
45	28	62.2	1487	2	US-08-760-489-2	Sequence 2, Appli
46	28	62.2	1487	2	US-08-760-489-4	Sequence 4, Appli
47	28	62.2	1622	4	US-08-972-927-6	Sequence 6, Appli
48	27	60.0	24	2	US-08-493-235-22	Sequence 22, Appl
49	27	60.0	26	2	US-08-493-235-21	Sequence 21, Appl
50	27	60.0	146	1	US-08-688-609-6	Sequence 6, Appli
51	27	60.0	146	3	US-09-002-832-6	Sequence 6, Appli
52	27	60.0	148	1	US-08-688-609-3	Sequence 3, Appli
53	27	60.0	148	3	US-09-002-832-3	Sequence 3, Appli
54	27	60.0	181	5	PCT-US93-07805-1	Sequence 1, Appli
55	27	60.0	220	4	US-09-004-731-95	Sequence 95, Appl
56	27	60.0	220	4	US-08-749-699-95	Sequence 95, Appl
57	27	60.0	226	4	US-09-004-731-8	Sequence 8, Appli
58	27	60.0	226	4	US-08-749-699-8	Sequence 8, Appli
59	27	60.0	240	1	US-08-488-113B-148	Sequence 148, App
60	27	60.0	240	1	US-08-477-484B-148	Sequence 148, App
61	27	60.0	240	2	US-08-646-360-148	Sequence 148, App
62	27	60.0	240	4	US-08-839-765-148	Sequence 148, App
63	27	60.0	240	4	US-09-136-389-148	Sequence 148, App
64	27	60.0	243	1	US-08-133-804-6	Sequence 6, Appli
65	27	60.0	243	1	US-08-461-838-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-071-710-36  
; Sequence 36, Application US/09071710  
; Patent No. 6130043  
; GENERAL INFORMATION:  
; APPLICANT: BILLING-MEDEL, PATRICIA  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GORDON, JULIAN  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: HODGES, STEVEN C.  
; APPLICANT: KRATOCHVIL, JON D.  
; APPLICANT: ROBERTS-RAPP, LISA  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STROUPE, STEPHEN D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; FOR DETECTING DISEASES OF THE PROSTATE  
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE PROSTATE  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible



OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ FOR Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICANT: US/09/071,710  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: 08/850,713  
FILING DATE: 02-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Becker, Cheryl L.  
REGISTRATION NUMBER: 35,441  
REFERENCE/DOCKET NUMBER: 6083.US.P1  
TELEPHONE: 847/935-1729  
TELEFAX: 847/938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 255 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 6130043e  
US-09-071-710-36

Query Match 100.0%; Score 45; DB 4; Length 255;  
Best Local Similarity 100.0%; Pred. No. 0.12; 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
DB 212 MGSIVQLSQS 221

RESULT 2  
US-08-459-287-2  
Sequence 2, Application US/08459287  
Patent No. 5686596  
GENERAL INFORMATION:  
APPLICANT: Mukherjee, Ranjan  
TITLE OF INVENTION: HUMAN PEROXISOME PROLIFERATOR ACTIVATED  
TITLE OF INVENTION: RECEPTOR  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM compatible  
OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,287  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/143,215  
FILING DATE: October 25, 1993  
APPLICATION NUMBER: 08/141,500  
FILING DATE: October 22, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 204/138  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440

TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 468 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-459-287-2

Query Match 68.9%; Score 31; DB 1; Length 468;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
DB 31 MGNIQEISQS 40

RESULT 3  
US-08-459-287-3  
Sequence 3, Application US/08459287  
Patent No. 5686596  
GENERAL INFORMATION:  
APPLICANT: Mukherjee, Ranjan  
TITLE OF INVENTION: HUMAN PEROXISOME PROLIFERATOR ACTIVATED  
TITLE OF INVENTION: RECEPTOR  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 611 West Sixth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM compatible  
OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,287  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/143,215  
FILING DATE: October 25, 1993  
APPLICATION NUMBER: 08/141,500  
FILING DATE: October 22, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 204/138  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 468 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-459-287-3

Query Match 68.9%; Score 31; DB 1; Length 468;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative .3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
DB 31 MGNIQEISQS 40

RESULT 4  
US-08-764-870-8  
; Sequence 8, Application US/08764870  
; Patent No. 6236946  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Thomas S  
; APPLICANT: Baxter, John D  
; APPLICANT: Fletcher, Robert J  
; APPLICANT: Wagner, Richard L  
; APPLICANT: Kushner, Peter J  
; APPLICANT: Aprelletti, James W  
; APPLICANT: West, Brian  
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand  
; TITLE OF INVENTION: Binding Domains  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooley Godward  
; STREET: Five Palo Alto Square, 3000 El Camino Real  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/764,870  
; FILING DATE: 13-DEC-1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/008,540  
; FILING DATE: 13-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/008,543  
; FILING DATE: 13-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/008,606  
; FILING DATE: 14-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Nakamura, Jackie N  
; REGISTRATION NUMBER: 35,966  
; REFERENCE/DOCKET NUMBER: UCAL-246/01US  
; TELEPHONE: (650)843-5000  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 468 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-764-870-8

Query Match 68.9%; Score 31; DB 4; Length 468;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
|||:|:|  
Db 31 MGNIQEISQS 40

RESULT 5  
US-08-800-198-8  
; Sequence 8, Application US/08800198  
; Patent No. 5942602  
; GENERAL INFORMATION:  
; APPLICANT: WELS, WINFRIED S.  
; APPLICANT: SCHMIDT, MATHIAS  
; APPLICANT: VAKALOPOULOU, EVANGELIA

; APPLICANT: SCHNEIDER, DOUGLAS  
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.  
; STREET: 2200 CLARENDON BLVD. SUITE 1400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: US  
; ZIP: 22201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/800,198  
; FILING DATE: 13-FEB-1997  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAMLET-KING, DIANA  
; REGISTRATION NUMBER: 33,302  
; REFERENCE/DOCKET NUMBER: SCH 1576  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-243-6333  
; TELEFAX: 703-243-6410  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
US-08-800-198-8

Query Match 64.4%; Score 29; DB 2; Length 240;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||:|:|  
Db 133 GSDIQLTQS 141

RESULT 6  
US-09-296-595-8  
; Sequence 8, Application US/09296595A  
; Patent No. 6129915  
; GENERAL INFORMATION:  
; APPLICANT: WELS, WINFRIED S.  
; APPLICANT: SCHMIDT, MATHIAS  
; APPLICANT: VAKALOPOULOU, EVANGELIA  
; APPLICANT: SCHNEIDER, DOUGLAS  
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES  
; FILE REFERENCE: SCH-1576 D1  
; CURRENT APPLICATION NUMBER: US/09/296,595A  
; CURRENT FILING DATE: 1999-04-23  
; EARLIER APPLICATION NUMBER: 08/800,198  
; EARLIER FILING DATE: 1997-02-13  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 240  
; TYPE: PRT  
; ORGANISM: Murine sp.  
US-09-296-595-8

Query Match 64.4%; Score 29; DB 4; Length 240;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||:||||  
Db 133 GSDIQLTQS 141.

RESULT 7  
US-08-235-838-5  
; Sequence 5, Application US/08235838  
; Patent No. 5571894  
; GENERAL INFORMATION:  
; APPLICANT: Wels, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5571894man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235,838  
FILING DATE: TBA  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828,832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 241 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-235-838-5

Query Match 64.4%; Score 29; DB 1; Length 241;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||:||||  
Db 134 GSDIQLTQS 142

RESULT 8  
US-08-235-838-11  
; Sequence 11, Application US/08235838  
; Patent No. 5571894  
; GENERAL INFORMATION:

; APPLICANT: Wels, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5571894man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235,838  
FILING DATE: TBA  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828,832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 241 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-235-838-11

Query Match 64.4%; Score 29; DB 1; Length 241;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||:||||  
Db 135 GSDIQLTQS 143

RESULT 9  
US-08-465-473B-5  
; Sequence 5, Application US/08465473B  
; Patent No. 5939531  
; GENERAL INFORMATION:  
; APPLICANT: Wels, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5939531man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NOVARTIS Corporation  
; STREET: 564 Morris Avenue

;  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901-6940  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,473B  
; FILING DATE: 5 June 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pfeiffer, Hesna J.  
; REGISTRATION NUMBER: 22,640  
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908)522 6940  
; TELEFAX: (908)522 6955  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 241 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-465-473B-5

Query Match 64.4%; Score 29; DB 2; Length 241;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||:|:|  
Db 134 GSDIQLTQS 142

RESULT 10  
US-08-465-473B-11  
; Sequence 11, Application US/08465473B  
; Patent No. 5939531  
; GENERAL INFORMATION:  
; APPLICANT: Wells, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5939531man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NOVARTIS Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901-6940  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,473B  
; FILING DATE: 5 June 1995

;  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pfeiffer, Hesna J.  
; REGISTRATION NUMBER: 22,640  
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908)522 6940  
; TELEFAX: (908)522 6955  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 241 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-465-473B-11

Query Match 64.4%; Score 29; DB 2; Length 241;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||:|:|  
Db 135 GSDIQLTQS 143

RESULT 11

US-08-887-352B-22  
; Sequence 22, Application US/08887352B  
; Patent No. 5994511  
; GENERAL INFORMATION:  
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe  
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of  
; TITLE OF INVENTION: Improving Polypeptides  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/887,352B  
; FILING DATE: 03-Jul-1997  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svoboda, Craig G.  
; REGISTRATION NUMBER: 39,044  
; REFERENCE/DOCKET NUMBER: P1123  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1489  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 248 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
US-08-887-352B-22

Query Match 64.4%; Score 29; DB 2; Length 248;



```
Best Local Similarity 66.7%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

QY 2 GSIVQLSQS 10
  11 :11:11
Db 135 GSDIQLTQS 143

RESULT 12
US-08-887-352B-23
; Sequence 23, Application US/08887352B
; Patent No. 5994511
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardiou, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
; TITLE OF INVENTION: Improving Polypeptides
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,352B
; FILING DATE: 03-Jul-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-887-352B-23

Query Match 64.4%; Score 29; DB 2; Length 248;
Best Local Similarity 66.7%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

QY 2 GSIVQLSQS 10
  11 :11:11
Db 135 GSDIQLTQS 143

RESULT 13
US-09-109-207C-22
; Sequence 22, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardiou, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypeptide
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; CURRENT FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/051,554
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 44
; SEQ ID NO 22
; LENGTH: 248
; TYPE: PRT
```

```
; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: Artificial
; LOCATION: 1-248
; OTHER INFORMATION: sfv sequence derived from MAE11
US-09-109-207C-22

Query Match 64.4%; Score 29; DB 4; Length 248;
Best Local Similarity 66.7%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

QY 2 GSIVQLSQS 10
  11 :11:11
Db 135 GSDIQLTQS 143

RESULT 14
US-09-109-207C-23
; Sequence 23, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardiou, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypept
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; CURRENT FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/051,554
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 44
; SEQ ID NO 23
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: Artificial
; LOCATION: 1-248
; OTHER INFORMATION: sfv sequence derived from MAE11
US-09-109-207C-23

Query Match 64.4%; Score 29; DB 4; Length 248;
Best Local Similarity 66.7%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

QY 2 GSIVQLSQS 10
  11 :11:11
Db 135 GSDIQLTQS 143

RESULT 15
US-08-797-689-18
; Sequence 18, Application US/08797689
; Patent No. 5876969
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Guitton, Jean-Dominique
; APPLICANT: Jung, Gerard
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,
; TITLE OF INVENTION: PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION
; TITLE OF INVENTION: CONTAINING SAID POLYPEPTIDES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
```

```

;
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.1
; SOFTWARE: Word 5.1 (PatentIn)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/797,689
; FILING DATE: 31-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,927
; FILING DATE: 28-JUL-1994
; APPLICATION NUMBER: FR 92/01064
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00085
; FILING DATE: 28-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith Ph.D., Julie K.
; REGISTRATION NUMBER: P-38,619
; REFERENCE/DOCKET NUMBER: ST92006-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 454-3839
; TELEFAX: (610) 454-3808
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-797-689-18

```

```

Query Match 64.4%; Score 29; DB 2; Length 249;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 GSIQVLSQS 10
   ||:||||
Db 141 GSIQVLSQS 149

```

```

RESULT 16
US-08-400-115-4
; Sequence 4, Application US/08400115
; Patent No. 5864019
; GENERAL INFORMATION:
; APPLICANT: KING, David John
; APPLICANT: MOUNTAIN, Andrew
; APPLICANT: OWENS, Raymond John
; APPLICANT: YARRANTON, Geoffrey Thomas
; TITLE OF INVENTION: Multivalent Antigen-Binding Proteins
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,115
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/127,136
; FILING DATE: 27-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/842,193
; FILING DATE: 17-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB91/00935

```

```

; FILING DATE: 11-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9012995.8
; FILING DATE: 11-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 040283/0106 CARA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 271 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-400-115-4

```

```

Query Match 64.4%; Score 29; DB 2; Length 271;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 GSIQVLSQS 10
   ||:||||
Db 144 GSIQVLSQS 152

```

```

RESULT 17
US-08-661-052-14
; Sequence 14, Application US/08661052
; Patent No. 5837243
; GENERAL INFORMATION:
; APPLICANT: Yashwant M. Deo
; APPLICANT: Joel Goldstein
; APPLICANT: Robert Graziano
; APPLICANT: Chezian Somasundaram
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED
; OF ANTI-PC RECEPTOR ANTIBODIES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/661,052
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/484,172
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MXI-043CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-08-661-052-14

Query Match 64.4%; Score 29; DB 2; Length 301;  
Best Local Similarity 66.7%; Pred. No. 2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
DB 153 GSDIQLTQS 161

RESULT 18

US-08-656-906-25  
; Sequence 25, Application US/08656906  
; Patent No. 5972901  
; GENERAL INFORMATION:  
; APPLICANT: Ferkol Jr., Thomas W.  
; APPLICANT: Davis, Pamela B.  
; APPLICANT: Ziady, Assem-Galal  
; TITLE OF INVENTION: Serpin Enzyme Complex Receptor -  
; TITLE OF INVENTION: Mediated Gene Transfer  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/656,906  
FILING DATE: 03-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION NUMBER: US 08/  
FILING DATE: 03-JUN-1996  
PRIOR APPLICATION DATA: WO WO 95/25809  
FILING DATE: 23-MAR-1995  
PRIOR APPLICATION NUMBER: US 08/216,534  
FILING DATE: 23-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: CASE-02280  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 301 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-656-906-25

Query Match 64.4%; Score 29; DB 2; Length 301;  
Best Local Similarity 77.8%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
DB 125 GSEVQLQQS 133

RESULT 19

US-09-217-847-25  
; Sequence 25, Application US/09217847  
; Patent No. 6200801  
; GENERAL INFORMATION:  
; APPLICANT: Ferkol Jr., Thomas W.  
; APPLICANT: Davis, Pamela B.  
; APPLICANT: Ziady, Assem-Galal  
; TITLE OF INVENTION: Serpin Enzyme Complex Receptor -  
; TITLE OF INVENTION: Mediated Gene Transfer  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/217,847  
FILING DATE: 09-09-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA: 08/656,906  
FILING DATE:  
PRIOR APPLICATION DATA: WO WO 95/25809  
FILING DATE: 23-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/216,534  
FILING DATE: 23-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: CASE-02280  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 301 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-217-847-25

Query Match 64.4%; Score 29; DB 4; Length 301;  
Best Local Similarity 77.8%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
DB 125 GSEVQLQQS 133

RESULT 20

US-08-875-811-41  
; Sequence 41, Application US/08875811  
; Patent No. 6045793  
; GENERAL INFORMATION:  
; APPLICANT: Rybak, Susanna M.  
; APPLICANT: Newton, Dianne L.  
; APPLICANT: Boque, Lluis  
; APPLICANT: Wlodawer, Alexander  
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,811
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/02588
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,800
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-244100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-875-811-41

Query Match 64.4%; Score 29; DB 3; Length 355;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
Db 121 GSEVQLQQS 129

RESULT 21
US-08-875-811-49
; Sequence 49, Application US/08875811
; Patent No. 6045793
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluis
; APPLICANT: Wlodawer, Alexander
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,811
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/02588
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,800
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-244100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-875-811-41
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/02588
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,800
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-244100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-875-811-49

Query Match 64.4%; Score 29; DB 3; Length 355;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
Db 121 GSEVQLQQS 129

RESULT 22
US-08-875-811-64
; Sequence 64, Application US/08875811
; Patent No. 6045793
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluis
; APPLICANT: Wlodawer, Alexander
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,811
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/02588
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,800
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-244100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
```



LENGTH: 355 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: protein  
LOCATION: 1..355  
OTHER INFORMATION: /note= "E6FB[Met-(-l)]SerrOnc"  
US-08-875-811-64

Query Match 64.4%; Score 29; DB 3; Length 355;  
Best Local Similarity 77.8%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10  
Db 121 GSEVQLQQS 129

RESULT 23  
US-08-875-811-45  
; Sequence 45, Application US/08875811  
; Patent No. 6045793  
; GENERAL INFORMATION:  
; APPLICANT: Rybak, Susanna M.  
; APPLICANT: Newton, Dianne L.  
; APPLICANT: Boque, Lluis  
; APPLICANT: Wlodawer, Alexander  
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/875,811  
; FILING DATE: 19-FEB-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US97/02588  
; FILING DATE: 19-FEB-1997  
; FILING DATE: 19-FEB-1997  
; APPLICATION NUMBER: US 60/011,800  
; FILING DATE: 21-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Faris, Susan K.  
; REGISTRATION NUMBER: 41,739  
; REFERENCE/DOCKET NUMBER: 015280-244100US  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 358 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-875-811-45

Query Match 64.4%; Score 29; DB 3; Length 358;  
Best Local Similarity 77.8%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10  
Db 239 GSEVQLQQS 247

RESULT 24  
US-08-875-811-51  
; Sequence 51, Application US/08875811  
; Patent No. 6045793  
; GENERAL INFORMATION:  
; APPLICANT: Rybak, Susanna M.  
; APPLICANT: Newton, Dianne L.  
; APPLICANT: Boque, Lluis  
; APPLICANT: Wlodawer, Alexander  
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/875,811  
; FILING DATE: 19-FEB-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US97/02588  
; FILING DATE: 19-FEB-1997  
; FILING DATE: 19-FEB-1997  
; APPLICATION NUMBER: US 60/011,800  
; FILING DATE: 21-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Faris, Susan K.  
; REGISTRATION NUMBER: 41,739  
; REFERENCE/DOCKET NUMBER: 015280-244100US  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 358 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-875-811-51

Query Match 64.4%; Score 29; DB 3; Length 358;  
Best Local Similarity 77.8%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10  
Db 239 GSEVQLQQS 247

RESULT 25  
US-08-875-811-47  
; Sequence 47, Application US/08875811  
; Patent No. 6045793  
; GENERAL INFORMATION:  
; APPLICANT: Rybak, Susanna M.  
; APPLICANT: Newton, Dianne L.  
; APPLICANT: Boque, Lluis  
; APPLICANT: Wlodawer, Alexander

;; TITLE OF INVENTION: Recombinant Ribonuclease Proteins  
;; NUMBER OF SEQUENCES: 64  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend and Crew LLP  
;; STREET: Two Embarcadero Center, Eighth Floor  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-3834  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/875,811  
;; FILING DATE: 19-FEB-1998  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: WO PCT/US97/02588  
;; FILING DATE: 19-FEB-1997  
;; APPLICATION NUMBER: US 60/011,800  
;; FILING DATE: 21-FEB-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Faris, Susan K.  
;; REGISTRATION NUMBER: 41,739  
;; REFERENCE/DOCKET NUMBER: 015280-244100US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 576-0200  
;; TELEFAX: (415) 576-0300  
;; INFORMATION FOR SEQ ID NO: 47:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 360 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-875-811-47

Query Match 64.4%; Score 29; DB 3; Length 360;  
Best Local Similarity 77.8%; Pred. No. 2.4e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSVQLSQS 10  
|| ||| ||  
Db 241 GSEVQLQQS 249

RESULT 26  
US-08-875-811-43  
;; Sequence 43, Application US/08875811  
;; Patent No. 6045793  
;; GENERAL INFORMATION:  
;; APPLICANT: Rybak, Susanna M.  
;; APPLICANT: Newton, Dianne L.  
;; APPLICANT: Boque, Lluís  
;; APPLICANT: Wlodawer, Alexander  
;; TITLE OF INVENTION: Recombinant Ribonuclease Proteins  
;; NUMBER OF SEQUENCES: 64  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend and Crew LLP  
;; STREET: Two Embarcadero Center, Eighth Floor  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-3834  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/875,811  
;; FILING DATE: 19-FEB-1998  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: WO PCT/US97/02588  
;; FILING DATE: 19-FEB-1997  
;; APPLICATION NUMBER: US 60/011,800  
;; FILING DATE: 21-FEB-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Faris, Susan K.  
;; REGISTRATION NUMBER: 41,739  
;; REFERENCE/DOCKET NUMBER: 015280-244100US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 576-0200  
;; TELEFAX: (415) 576-0300  
;; INFORMATION FOR SEQ ID NO: 43:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 379 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-875-811-43

Query Match 64.4%; Score 29; DB 3; Length 379;  
Best Local Similarity 77.8%; Pred. No. 2.5e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSVQLSQS 10  
|| ||| ||  
Db 263 GSEVQLQQS 271

RESULT 27  
US-08-661-052-16  
;; Sequence 16, Application US/08661052  
;; Patent No. 5837243  
;; GENERAL INFORMATION:  
;; APPLICANT: Yashwant M. Deo  
;; APPLICANT: Joel Goldstein  
;; APPLICANT: Robert Graziano  
;; APPLICANT: Chezian Somasundaram  
;; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED  
;; OF ANTI-FC RECEPTOR ANTIBODIES  
;; NUMBER OF SEQUENCES: 16  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: LAHIVE & COCKFIELD  
;; STREET: 60 State Street, Suite 510  
;; CITY: Boston  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02109-1875  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/661,052  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/484,172  
;; FILING DATE: 07-JUNE-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Arnold, Beth E.  
;; REGISTRATION NUMBER: 35,430  
;; REFERENCE/DOCKET NUMBER: MXI-043CP  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617)227-7400  
;; TELEFAX: (617)227-5941  
;; INFORMATION FOR SEQ ID NO: 16:  
;; SEQUENCE CHARACTERISTICS:

; LENGTH: 553 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-661-052-16

Query Match 64.4%; Score 29; DB 2; Length 553;  
Best Local Similarity 66.7%; Pred. No. 3.7e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
Db 153 GSDIQLTQS 161

RESULT 28  
US-08-235-838-14  
; Sequence 14, Application US/082335838  
; Patent No. 5571894  
; GENERAL INFORMATION:  
; APPLICANT: Wells, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5571894man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/235,838  
; FILING DATE: TBA  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
; TELEPHONE: (919)541-8614  
; TELEFAX: (919)541-8689  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 637 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-235-838-14

Query Match 64.4%; Score 29; DB 1; Length 637;  
Best Local Similarity 66.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

Db 165 GSDIQLTQS 173

RESULT 29  
US-08-235-838-16  
; Sequence 16, Application US/082335838  
; Patent No. 5571894  
; GENERAL INFORMATION:  
; APPLICANT: Wells, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria  
; APPLICANT: Groner, Bernd  
; APPLICANT: Hardman, No. 5571894man  
; APPLICANT: Zwickl, Markus  
; TITLE OF INVENTION: Recombinant Antibodies Specific for a  
; TITLE OF INVENTION: Growth Factor Receptor  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/235,838  
; FILING DATE: TBA  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/828,832  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 91-810079.3  
; FILING DATE: 05-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
; TELEPHONE: (919)541-8614  
; TELEFAX: (919)541-8689  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 637 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-235-838-16

Query Match 64.4%; Score 29; DB 1; Length 637;  
Best Local Similarity 66.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
Db 166 GSDIQLTQS 174

RESULT 30  
US-08-465-473B-14  
; Sequence 14, Application US/08465473B  
; Patent No. 5939531  
; GENERAL INFORMATION:  
; APPLICANT: Wells, Winfried S.  
; APPLICANT: Hynes, Nancy E.  
; APPLICANT: Harwerth, Ina-Maria

APPLICANT: Groner, Bernd  
APPLICANT: Hardman, No. 5939531man  
APPLICANT: Zwickl, Markus  
TITLE OF INVENTION: Recombinant Antibodies Specific for a  
TITLE OF INVENTION: Growth Factor Receptor  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NOVARTIS Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901-6940  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,473B  
FILING DATE: 5 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828,832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,473B  
FILING DATE: 5 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828,832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pfeiffer, Hesna J.  
REGISTRATION NUMBER: 22,640  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)522 6940  
TELEFAX: (908)522 6955  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-465-473B-14

Query Match 64.4%; Score 29; DB 2; Length 637;  
Best Local Similarity 66.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
|| :||:||  
Db 165 GSDIQLTQS 173

RESULT 31  
US-08-465-473B-16  
Sequence 16, Application US/08465473B  
Patent No. 5939531  
GENERAL INFORMATION:  
APPLICANT: Wels, Winfried S.  
APPLICANT: Hynes, Nancy E.  
APPLICANT: Harwerth, Ina-Maria  
APPLICANT: Groner, Bernd  
APPLICANT: Hardman, No. 5939531man  
TITLE OF INVENTION: Recombinant Antibodies Specific for a  
TITLE OF INVENTION: Growth Factor Receptor  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NOVARTIS Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA

ZIP: 07901-6940  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,473B  
FILING DATE: 5 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828,832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pfeiffer, Hesna J.  
REGISTRATION NUMBER: 22,640  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)522 6940  
TELEFAX: (908)522 6955  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-465-473B-16

Query Match 64.4%; Score 29; DB 2; Length 637;  
Best Local Similarity 66.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
|| :||:||  
Db 166 GSDIQLTQS 174

RESULT 32  
US-08-235-838-7  
Sequence 7, Application US/08235838  
Patent No. 5571894  
GENERAL INFORMATION:  
APPLICANT: Wels, Winfried S.  
APPLICANT: Hynes, Nancy E.  
APPLICANT: Harwerth, Ina-Maria  
APPLICANT: Groner, Bernd  
APPLICANT: Hardman, No. 5571894man  
TITLE OF INVENTION: Recombinant Antibodies Specific for a  
TITLE OF INVENTION: Growth Factor Receptor  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235,838  
FILING DATE: TBA  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828,832



```

; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-235-838-7
```

```

Query Match 64.4%; Score 29; DB 1; Length 711;
Best Local Similarity 66.7%; Pred. NO. 4.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```

OY 2 GSIVQLSQS 10
Db 155 GSDIQLTQS 163
```

```

RESULT 33
US-08-465-473B-7
; Sequence 7, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hgsna J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 7:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-473B-7

Query Match 64.4%; Score 29; DB 2; Length 711;
Best Local Similarity 66.7%; Pred. NO. 4.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
Db 155 GSDIQLTQS 163

RESULT 34
US-08-633-770A-2
; Sequence 2, Application US/08633770A
; Patent No. 5908760
; GENERAL INFORMATION:
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; TITLE OF INVENTION: ALPHA-1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: OCT-15-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DYOU6.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-633-770A-2

Query Match 64.4%; Score 29; DB 2; Length 1070;
Best Local Similarity 66.7%; Pred. NO. 7.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
Db 1020 GSIVDVSQL 1028

```

RESULT 35  
US-08-902-486-13  
; Sequence 13, Application US/08902486  
; Patent No. 6140075  
; GENERAL INFORMATION:  
; APPLICANT: Russel, David R.  
; APPLICANT: Fuller, James T.  
; TITLE OF INVENTION: METHOD FOR PRODUCING ANTIBODIES AND  
; TITLE OF INVENTION: PROTEIN TOXINS IN PLANT CELLS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/902,486  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 670513.90261  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 241 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-902-486-13

Query Match 62.2%; Score 28; DB 4; Length 241;  
Best Local Similarity 66.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||:||||  
Db 128 GSDIQLQSQS 136  
  
RESULT 36  
US-08-748-506-10  
; Sequence 10, Application US/08748506  
; Patent No. 6159707  
; GENERAL INFORMATION:  
; APPLICANT: Ronnett et al.  
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/748,506  
; FILING DATE: 08-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/033,751  
; FILING DATE: 09-NOV-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 74940  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5600  
; TELEFAX: 312-616-5700  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 321 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-748-506-10

Query Match 62.2%; Score 28; DB 4; Length 321;  
Best Local Similarity 60.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
||:||||  
Db 157 MGCIVSLGQT 166  
  
RESULT 37  
US-08-748-506-18  
; Sequence 18, Application US/08748506  
; Patent No. 6159707.  
; GENERAL INFORMATION:  
; APPLICANT: Ronnett et al.  
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/748,506  
; FILING DATE: 08-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/033,751  
; FILING DATE: 09-NOV-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 74940  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5600  
; TELEFAX: 312-616-5700  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 321 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-748-506-18

Query Match 62.2%; Score 28; DB 4; Length 321;  
Best Local Similarity 60.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
DB 157 MGCIVSIGQT 166

RESULT 38  
US-08-902-486-15  
; Sequence 15, Application US/08902486  
; Patent No. 6140075  
; GENERAL INFORMATION:  
; APPLICANT: Russel, David R.  
; APPLICANT: Fuller, James T.  
; TITLE OF INVENTION: METHOD FOR PRODUCING ANTIBODIES AND  
; TITLE OF INVENTION: PROTEIN TOXINS IN PLANT CELLS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/902,486  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 670513.90261  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 496 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-902-486-15

Query Match 62.2%; Score 28; DB 4; Length 496;  
Best Local Similarity 66.7%; Pred. No. 5.2e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIIVQLSQS 10  
DB 383 GSDIQLQCS 391

RESULT 39  
US-08-933-750C-36  
; Sequence 36, Application US/08933750C  
; Patent No. 5932442  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750C  
; FILING DATE: September 23, 1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 612 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BRAINON01  
; CLONE: 1832295  
US-08-933-750C-36

Query Match 62.2%; Score 28; DB 2; Length 612;  
Best Local Similarity 77.8%; Pred. No. 6.5e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
DB 483 GSAGQLSQS 491

RESULT 40  
US-09-234-613-36  
; Sequence 36, Application US/09234613  
; Patent No. 6132973  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/234,613  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/933,750  
FILING DATE: September 23, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 612 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAINON01  
CLONE: 1832295  
US-09-234-613-36

Query Match 62.2%; Score 28; DB 4; Length 612;  
Best Local Similarity 77.8%; Pred. No. 6.5e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
|| |||||  
Db 483 GSAGQLSQS 491

RESULT 41  
US-07-674-287B-2  
Sequence 2, Application US/07674287B  
Patent No. 5414076  
GENERAL INFORMATION:  
APPLICANT: Bryan Mark O'Hara  
TITLE OF INVENTION: Gibbon Ape Leukemia  
TITLE OF INVENTION: Virus Receptor  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dr. Karen A. Lowney  
ADDRESSEE: American Cyanamid Company  
STREET: 1937 West Main Street  
STREET: P.O. Box 60  
CITY: Stamford  
STATE: CT  
COUNTRY: USA  
ZIP: 06904-0060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC AT  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII converted from IBM DW4  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/674,287B  
FILING DATE: 19910325  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lowney, Karen A., Dr.

REGISTRATION NUMBER: 31,274  
REFERENCE/DOCKET NUMBER: 31,104-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 203 321 2361  
TELEFAX: 203 321 2971  
TELEX: 710 474 4059  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 680 Amino Acid Residues  
TYPE: AMINO ACID SEQUENCE  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Protein  
US-07-674-287B-2

Query Match 62.2%; Score 28; DB 1; Length 680;  
Best Local Similarity 55.6%; Pred. No. 7.3e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
|:|:|:|:  
Db 357 GNLVQFSQA 365

RESULT 42  
US-08-436-900A-2  
Sequence 2, Application US/08436900A  
Patent No. 5874264  
GENERAL INFORMATION:  
APPLICANT: O'Hara, Bryan M.  
TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Home Products  
STREET: One Campus Drive  
CITY: Parsippany  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07054  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,900A  
FILING DATE: 08-MAY-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 31,104-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-683-2158  
TELEFAX: 201-683-4117  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 680 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-436-900A-2

Query Match 62.2%; Score 28; DB 2; Length 680;  
Best Local Similarity 55.6%; Pred. No. 7.3e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
|:|:|:|:  
Db 357 GNLVQFSQA 365



```

RESULT 43
US-08-436-900A-4
; Sequence 4, Application US/08436900A
; Patent No. 5874264
; GENERAL INFORMATION:
; APPLICANT: O'Hara, Bryan M.
; TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,900A
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnhard, Elizabeth M.
; REGISTRATION NUMBER: 31,088
; REFERENCE/DOCKET NUMBER: 31,104-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-683-2158
; TELEFAX: 201-683-4117
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-436-900A-4

Query Match 62.2%; Score 28; DB 2; Length 682;
Best Local Similarity 55.6%; Pred. No. 7.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
Db 360 GNLVQFSQT 368

RESULT 44
US-08-460-269C-2
; Sequence 2, Application US/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 910 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2

Query Match 62.2%; Score 28; DB 4; Length 910;
Best Local Similarity 66.7%; Pred. No. 9.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
Db 302 GSSVELAQs 310

RESULT 45
US-08-760-489-2
; Sequence 2, Application US/08760489
; Patent No. 5830696
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,489
; FILING DATE: 05-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/008,311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-760-489-2

```

Query Match 62.2%; Score 28; DB 2; Length 1487;  
Best Local Similarity 44.4%; Pred. NO. 1.6e+03;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:|||||  
Db 221 LGSVIKLGQ 229

Search completed: June 28, 2001, 11:56:15  
Job time: 196 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:54:17 ; Search time 66.55 Seconds  
(without alignments)  
9.110 Million cell updates/sec

Title: US-09-439-313-573  
Perfect score: 45  
Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	231	AA54369	Amino acid sequenc
2	45	100.0	255	AAW85068	Protein encoded by
3	45	100.0	255	AAW29268	Human prostate-rel
4	45	100.0	553	AAW71869	Amino acid encoded
5	45	100.0	553	AAW69385	Prostate tumour sp
6	45	100.0	553	AAW28527	Protein encoded by
7	45	100.0	553	AAW82002	Human immunogenic
8	33	73.3	3413	AAW52849	A. mediterranei ri
9	31	68.9	468	AAW74053	Human peroxisome p
10	31	68.9	468	AAW21633	Ligand binding dom
11	31	68.9	468	AAW05470	Human PPAR-alpha p

12	31	68.9	468	22	AAW20342	Peroxisome prolife
13	30	66.7	288	21	AAG11352	Arabidopsis thalia
14	30	66.7	288	21	AAG16074	Arabidopsis thalia
15	30	66.7	303	21	AAG11351	Arabidopsis thalia
16	30	66.7	303	21	AAG16073	Arabidopsis thalia
17	30	66.7	425	21	AAG11350	Arabidopsis thalia
18	30	66.7	425	21	AAG16072	Arabidopsis thalia
19	30	66.7	750	21	AAW26599	Synechocystis sp p
20	30	66.7	2559	20	AAW41012	Amino acid sequenc
21	30	66.7	5069	19	AAW52846	A. mediterranei ri
22	29	64.4	36	22	AAW65054	Gene #7 associated
23	29	64.4	155	18	AAW55323	H. pylori ORF hp2e
24	29	64.4	160	20	AAW60166	Human endometrium
25	29	64.4	230	21	AAW42674	Human ORFX ORF2438
26	29	64.4	240	16	AAW85494	ErBB-2-specific sc
27	29	64.4	240	16	AAW85495	ScFv(FWP51). Synt
28	29	64.4	240	17	AAW05133	Single chain antib
29	29	64.4	240	17	AAW05134	Single chain antib
30	29	64.4	240	18	AAW25784	Anti-gp130 antibod
31	29	64.4	240	19	AAW71243	scFv comprising he
32	29	64.4	241	13	AAW26979	Fv heavy chain/lig
33	29	64.4	241	13	AAW26981	FWP51 fusion prote
34	29	64.4	242	22	AAW31426	Protein used for i
35	29	64.4	248	16	AAW77606	Murine 5G1.1M1 scF
36	29	64.4	248	18	AAW01829	scFv(1.1ASML) gene
37	29	64.4	248	20	AAW95668	Mus musculus anti-
38	29	64.4	248	20	AAW95667	Mus musculus anti-
39	29	64.4	248	22	AAW76956	SFV fragment of e2
40	29	64.4	248	22	AAW76957	SFV fragment of e2
41	29	64.4	251	9	AAW80152	Multifunctional pr
42	29	64.4	253	21	AAW28159	Murine anti-EGP-2
43	29	64.4	271	13	AAW20185	Sequence of the sh
44	29	64.4	272	14	AAW34672	26-10 sfv. Synthe
45	29	64.4	272	17	AAW00557	Nematode salivary
46	29	64.4	272	19	AAW43913	Mus musculus antib
47	29	64.4	284	21	AAW38451	Fragment of human
48	29	64.4	285	18	AAW55647	H. pylori ORF 06ep
49	29	64.4	285	18	AAW55573	H. pylori ORF 06ep
50	29	64.4	287	19	AAW98675	H. pylori GHPO 548
51	29	64.4	288	20	AAW82743	Fusion protein PNG
52	29	64.4	289	22	AAW65048	Gene #7 associated
53	29	64.4	301	18	AAW11507	Single chain, huma
54	29	64.4	301	19	AAW37085	Anti-human SC sing
55	29	64.4	301	20	AAW73217	Multispecific sing
56	29	64.4	301	22	AAW61959	Single chain human
57	29	64.4	305	21	AAW59264	Antibody 4H5 H cha
58	29	64.4	305	21	AAW51141	Murine derived pro
59	29	64.4	331	20	AAW94268	H6FXtripBscFv(CEA6
60	29	64.4	347	22	AAW64999	Human secreted pro
61	29	64.4	355	18	AAW35129	R. pipiens recombi
62	29	64.4	355	18	AAW35125	R. pipiens recombi
63	29	64.4	358	18	AAW35127	R. pipiens recombi
64	29	64.4	358	18	AAW35130	R. pipiens recombi
65	29	64.4	360	18	AAW35128	R. pipiens recombi

ALIGNMENTS

RESULT	1
AA54369	
ID	AA54369 standard; Protein; 231 AA.
XX	
AC	AA54369;
XX	
DT	06-APR-2000 (first entry)
XX	
DE	Amino acid sequence of a prostate cancer-associated protein.
XX	
KW	Prostate cancer-associated gene; Incyte clone 1864683; bone cancer;
KW	cell proliferation; cancer; adrenal gland cancer; bladder cancer;
KW	prostate cancer.
XX	

OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..47  
FT /note= "signal peptide"  
FT Modified-site 100  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 142  
FT /note= "potential casein kinase II phosphorylation site"  
FT Modified-site 147  
FT /note= "potential protein kinase C phosphorylation site"  
FT Domain 188..209  
FT /note= "potential transmembrane domain"  
XX  
XX WO9967384-A2.  
XX  
XX 29-DEC-1999.  
XX  
XX 15-JUN-1999; 99WO-US13524.  
XX  
XX 22-JUN-1998; 98US-0102615.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Walker MG, Volkmuth W, Klingler TM, Sprinzak EA;  
XX N-PSDB; AAZ45677.  
XX  
XX WPI: 2000-126631/11.  
XX N-PSDB; AAZ45677.  
XX  
XX Identifying biomolecules for the diagnosis and treatment of diseases  
XX associated with cell-proliferation -  
XX  
XX Claim 10; Page 50-51; 52pp; English.  
XX  
XX The present sequence is encoded by a prostate cancer-associated gene,  
XX whose cDNA is represented by incyte clone number 1864683. The cDNA  
XX describes a method for identifying biomolecules for the diagnosis or  
XX treatment of diseases associated with cell proliferation. The method  
XX comprises examining polynucleotides, consisting of prostate cancer-  
XX specific genes, and genes of unknown function, expressed in cDNA  
XX libraries. The patterns of both gene sets are compared to identify  
XX genes of unknown function with similar expression patterns to the  
XX prostate cancer-specific genes. The biomolecules identified by the  
XX method form pharmaceutical compositions useful for the diagnosis and  
XX treatment of diseases associated with cell proliferation. Such diseases  
XX include cancer of the adrenal gland, bladder and bone, but especially  
XX prostate cancer. The method may also be applied using other disease-  
XX specific genes. The prostate cancer-specific genes facilitate the  
XX diagnosis and treatment of cell proliferation disorders.  
XX  
XX Sequence 231 AA;  
  
Query Match 100.0%; Score 45; DB 21; Length 231;  
Best Local Similarity 100.0%; Pred. No. 0.075; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;  
  
QY 1 MGSIVQLSQS 10  
Db 188 mgsivqlsq 197  
  
RESULT 2  
AAW85068  
ID AAW85068 standard; Protein; 255 AA.  
XX  
XX AAW85068;  
AC  
XX 12-FEB-1999 (first entry)  
DT  
XX Protein encoded by the consensus sequence of the pS108 gene.  
DE  
XX

KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
XX drug screening; gene therapy.  
XX Homo sapiens.  
XX WO9850567-A1.  
XX  
XX 12-NOV-1998.  
XX  
XX 01-MAY-1998; 98WO-US08930.  
XX  
XX 02-MAY-1997; 97US-0850713.  
XX (ABBO ) ABBOTT LAB.  
XX  
XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;  
XX Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;  
XX Roberts-Rapp L, Russell JC, Stroupe SD;  
XX  
XX WPI: 1999-034731/03.  
XX N-PSDB; AAV71181.  
XX  
XX New isolated prostate-specific polynucleotides - used to develop  
XX products for the diagnosis and treatment of prostate diseases, e.g.  
XX benign hyperplasia, prostatic or prostate cancer  
XX  
XX Claim 17; Pages 99-100; 122pp; English.  
XX  
XX The present sequence is encoded by the consensus sequence for a pS108  
XX gene. The sequence is derived from overlapping clones AAV71166-79. The  
XX clone sequences are pS108 gene-specific. They are used in the method  
XX of the invention. The specification describes a method for detecting the  
XX presence of a target pS108 polynucleotide in a test sample. The method  
XX comprises contacting the test sample with at least 1 pS108-specific  
XX polynucleotide or complement, and detecting the presence of the target  
XX pS108 polynucleotide. The products can be used for detecting,  
XX diagnosing, staging, monitoring, prognosticating, in vivo imaging,  
XX preventing or treating, or determining predisposition to diseases or  
XX conditions of the prostate such as benign prostatic hyperplasia (BPH),  
XX prostatitis, prostatic intraepithelial neoplasia (PIN) and cancer. In  
XX particular the products can be used in drug screening and gene therapy.  
XX  
XX Sequence 255 AA;  
  
Query Match 100.0%; Score 45; DB 20; Length 255;  
Best Local Similarity 100.0%; Pred. No. 0.084; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;  
  
QY 1 MGSIVQLSQS 10  
Db 212 mgsivqlsq 221  
  
RESULT 3  
AAB29268  
ID AAB29268 standard; Protein; 255 AA.  
XX  
XX AAB29268;  
AC  
XX 07-FEB-2001 (first entry)  
DT  
XX  
XX Human prostate-related pS108 partial protein sequence SEQ ID NO: 36.  
XX  
XX Human; prostate cancer; pS108; antibody; tumour; metastasis.  
XX  
XX Homo sapiens.  
XX  
XX US6130043-A.  
XX  
XX 10-OCT-2000.



PF 01-MAY-1998; 98US-0071710.  
XX  
PR 02-MAY-1997; 97US-0850713.  
XX  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Cohen M, Colpitts TL, Friedman PN, Gordon J, Granados EN;  
PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;  
PI Kratochvil JD, Russell JC, Hodges SC;  
XX  
DR WPI; 2000-655655/63.  
XX  
PT Methods for detecting target prostate-specific polynucleotides or  
PT diseases of the prostate (e.g. prostate cancer), comprising detecting  
PT the presence of any of PSl08 nucleic acid sequences in a test sample -  
XX  
PS Example 1; Column 85-88; 55pp; English.  
XX  
CC The present invention is related to a number of partial coding and  
CC protein sequences for the human prostate tissue protein PSl08. These  
CC sequences can be used in the diagnosis and prognosis of prostate  
CC diseases, particularly prostate cancer. They can also be used to produce  
CC antibodies which can be used in treatment. The present sequence is one  
CC of the PSl08 partial protein sequences.  
XX  
SQ Sequence 255 AA;

Query Match 100.0%; Score 45; DB 21; Length 255;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 212 mgsivqlsqsq 221  
IIIIIIIIII

RESULT 4  
AAW71869  
ID AAW71869 standard; Protein; 553 AA.  
XX  
AC AAW71869;  
XX  
DT 06-JAN-1999 (first entry)  
XX  
DE Amino acid encoded by prostate tumour clone L1-12.  
XX  
KW Prostate; cancer; tumour; vaccine; immunogen; clone.  
XX  
OS Homo sapiens.  
XX  
PN WO9837093-A2.  
XX  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03492.  
XX  
PR 09-FEB-1998; 98US-0020956.  
PR 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
DR WPI; 1998-609886/51.  
DR N-PSDB; AAV61201.  
XX  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
XX  
PS Example 1; Page 82-84; 130pp; English.  
XX

CC The present sequence is an immunogenic portion of a prostate tumour  
CC protein. The immunogen, or the DNA encoding it, can be used as a  
CC vaccine for the treatment of prostate cancer. The immunogen was  
CC isolated from a prostate tumour cDNA library obtained by subtracting  
CC a prostate tumour cDNA expression library with a normal tissue cDNA  
CC library.  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 45; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 510 mgsivqlsqsq 519  
IIIIIIIIII

RESULT 5  
AAW69385  
ID AAW69385 standard; Protein; 553 AA.  
XX  
AC AAW69385;  
XX  
DT 08-DEC-1998 (first entry)  
XX  
DE Prostate tumour specific gene clone L1-12 protein.  
XX  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
KW therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO9837418-A2.  
XX  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03690.  
XX  
PR 09-FEB-1998; 98US-0904809.  
PR 25-FEB-1997; 97US-0806596.  
PR 01-AUG-1997; 97US-0904809.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Dillon DC, Xu J;  
XX  
DR WPI; 1998-480805/41.  
DR N-PSDB; AAV58586.  
XX  
PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
XX  
PS Example 1; Page 87-89; 141pp; English.  
XX  
CC This sequence is encoded by a human prostate tumour specific gene, and  
CC can be used in the method of the invention. The method is for detecting  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC this protein sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 45; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 510 mgsivqlsqsq 519

RESULT 6  
AAB28527  
ID AAB28527 standard; Protein; 553 AA.  
AC AAB28527;  
XX  
DT 07-FEB-2001 (first entry)  
DE Protein encoded by human breast tumour cDNA clone P501S.  
DE Human; breast tumour antigen; cytostatic; immunotherapy;  
KW breast cancer; vaccine.  
XX Homo sapiens.  
OS  
XX WO2000061756-A2.  
PN  
XX 19-OCT-2000.  
PD  
XX 10-APR-2000; 2000WO-US09688.  
PF  
XX 09-APR-1999; 99US-0288950.  
PR 02-JUL-1999; 99US-0346327.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Reed SG, Xu J, Dillon DC;  
PI  
XX WPI; 2000-638568/61.  
DR N-PSDB; AAC79473.  
XX  
XX A novel isolated polypeptide comprising an immunogenic portion of a  
PT breast cancer protein useful in the detection and treatment of breast  
PT cancer -  
PT  
XX  
PS Claim 2; Page 92-93; 95pp; English.

XX The present sequence is encoded by a cDNA sequence which was isolated  
CC from a breast tumour cDNA library. It is provided in a specification  
CC relating to compounds for immunotherapy and diagnosis of breast cancer.  
CC Breast tumour antigens and the polynucleotides that encode them may be  
CC used in the production of a pharmaceutical composition to be used in the  
CC treatment of breast cancer. Proliferated T cells and incubated antigen  
CC presenting cells are also required. The polypeptides and polynucleotides  
CC may also be used to produce a vaccine.  
XX  
XX Sequence 553 AA;

Query Match 100.0%; Score 45; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 510 mgsivqlsqsq 519

RESULT 7  
AAY82002  
ID AAY82002 standard; Protein; 553 AA.  
XX AAY82002;  
AC  
XX 13-JUN-2000 (first entry)  
DT  
XX Human immunogenic prostate tumour protein sequence SEQ ID NO:113.  
DE  
XX

KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
XX immunogenic; cytostatic; vaccine.  
OS Homo sapiens.  
XX WO200004149-A2.  
PN  
XX 27-JAN-2000.  
PD  
XX 14-JUL-1999; 99WO-US15838.  
PF  
XX 14-JUL-1998; 98US-0115453.  
PR 14-JUL-1998; 98US-0116134.  
PR 23-SEP-1998; 98US-0159812.  
PR 23-SEP-1998; 98US-0159822.  
PR 15-JAN-1999; 99US-0232149.  
PR 15-JAN-1999; 99US-0232880.  
PR 09-APR-1999; 99US-0288946.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;  
PI  
XX WPI; 2000-171268/15.  
DR  
XX New polypeptide useful for treating and diagnosing prostate cancer  
PT comprises an immunogenic portion of prostate tumor protein -  
PT  
XX  
PS Claim 3; Page 138-139; 263pp; English.

XX The present invention describes isolated polypeptides, comprising an  
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides  
CC and polynucleotides encoding them have cytostatic activity and can be  
CC used in vaccines and in gene therapy. The polypeptides and  
CC polynucleotides encoding them, antigen presenting cells which express  
CC the polypeptides, antibodies against the polypeptides and vaccines  
CC comprising them can be used for inhibiting the development of prostate  
CC cancer in a patient. The polypeptides can be used to generate antibodies  
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of  
CC the polynucleotides encoding the polypeptides can be used as a probe or  
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and  
CC AAY82000 to AAY82020 represent sequences used in the exemplification of  
CC the present invention.

XX Sequence 553 AA;

Query Match 100.0%; Score 45; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 510 mgsivqlsqsq 519

RESULT 8  
AAW52849  
ID AAW52849 standard; Protein; 3413 AA.  
XX  
XX AAW52849;  
AC  
XX 24-JUL-1998 (first entry)  
DT  
XX A. mediterranei rifamycin synthesis gene cluster fragment protein E.  
DE  
XX Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
KW polyketide synthase; actinomycete; ansamycin.  
XX  
XX Amycolatopsis mediterranei.  
OS  
XX WO9807868-A1.  
PN  
XX

PD 26-FEB-1998.  
XX  
PF 18-AUG-1997; 97WO-EP04495.  
XX  
PR 20-AUG-1996; 96EP-0810551.  
XX  
PA (NOVS ) NOVARTIS AG.  
XX  
PI Engel N, Schupp T, Toupet C;  
XX  
DR WPI; 1998-169172/15.  
DR N-PSDB; AAV21187.  
XX  
PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used  
PT to produce rifamycin and rifamycin analogues  
XX  
PS Claim 6; Page 170-187; 205pp; English.  
XX  
CC The present sequence represents a Amycolatopsis mediterranei rifamycin  
CC synthesis gene cluster ORF E protein from the present invention. The  
CC DNA fragment comprises a DNA region involved directly or indirectly  
CC in the gene cluster responsible for rifamycin synthesis, including  
CC the adjacent DNA regions to the right and left which, by reason of  
CC their function in connection with rifamycin biosynthesis, qualify  
CC as constituents of this rifamycin gene cluster, and functional  
CC fragments, derivatives or constituents of these. The Amycolatopsis  
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used  
CC for producing rifamycin, rifamycin analogues or precursors. It can also  
CC be used for inactivating or modifying genes involved in ansamycin or  
CC rifamycin biosynthesis. The DNA can be used for constructing mutant  
CC actinomycetes strains from which the natural rifamycin or ansamycin  
CC biosynthesis gene cluster has been partly or completely deleted. The  
CC DNA fragment can be used for assembling a library of polyketide  
CC synthases, which can be used for assembling a library of polyketides.  
CC A hybridisation probe of the invention can be used for identifying DNA  
CC fragments involved in the biosynthesis of ansamycins.  
XX  
SQ Sequence 3413 AA;

Query Match 73.3%; Score 33; DB 19; Length 3413;  
Best Local Similarity 60.0%; Pred. No. 5.4e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 2588 lgavvqlpqs 2597  
:|::||| ||

RESULT 9  
AAR74053  
ID AAR74053 standard; Protein; 468 AA.  
XX  
AC AAR74053;  
XX  
DT 24-OCT-1995 (first entry)  
XX  
DE Human peroxisome proliferator activated receptor.  
XX  
KW PPAR; hyperlipidaemia; atherosclerosis.  
XX  
OS Homo sapiens.  
XX  
PN WO9511974-A.  
XX  
PD 04-MAY-1995.  
XX  
PF 19-OCT-1994; 94WO-US11897.  
XX  
PR 26-OCT-1993; 93US-0143215.  
PR 22-OCT-1993; 93US-0141500.  
XX  
PA (LIGA-) LIGAND PHARM INC.

XX Mukherjee R;  
PI  
XX  
DR WPI; 1995-178867/23.  
DR N-PSDB; AAQ92081.  
XX  
PT New human peroxisome proliferator activated receptor DNA - used  
PT to screen for agents for the treatment of hyperlipidaemia and  
PT diseases such as atherosclerosis  
XX  
PS Claim 3; Fig 1; 19pp; English.  
XX  
CC The sequence is that of the human peroxisome proliferator  
CC activated receptor (PPAR) isolated from a human liver cDNA library  
CC using a probe based on rat PPAR DNA. The human PPAR can be used  
CC for screening cpds. for improved pharmacological profiles for the  
CC treatment of hyperlipidaemia with higher potency, efficacy and fewer  
CC side effects. It can be used to screen cpds. active as primary  
CC endogenous inducers of the human PPAR. Also PPAR can be used for  
CC establishing the tissue specific expression pattern of human PPAR to  
CC aid treatment of diseases such as atherosclerosis.  
XX  
SQ Sequence 468 AA;

Query Match 68.9%; Score 31; DB 16; Length 468;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 31 mgniqeaisqs 40  
||:| :|||

RESULT 10  
AAV21633  
ID AAY21633 standard; protein; 468 AA.  
XX  
AC AAY21633;  
XX  
DT 11-AUG-1999 (first entry)  
XX  
DE Ligand binding domain of nuclear receptor hPPARalpha.  
XX  
KW Thyroid hormone receptor; aromatic compound; ligand binding domain;  
KW alpha-glycerophosphate dehydrogenase; cardiac; obesity; triglyceride;  
KW plasma cholesterol; anti-hypertriglyceridaemic; atherosclerosis; GPDH;  
KW thyroid hormone replacement therapy; nuclear receptor.  
XX  
OS Homo sapiens.  
XX  
PN WO9926966-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 25-NOV-1998; 98WO-US25296.  
XX  
PR 26-NOV-1997; 97US-0980115.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Apriletti JW, Baxter JD, Fletterick RJ, Kushner PJ;  
PI Scanlan TS, Shiau AK, Wagner RL, West BL;  
XX  
DR WPI; 1999-357810/30.  
XX  
PT Modulating activity of a thyroid hormone receptor  
XX  
PS Disclosure; Fig 3I-R; 447pp; English.  
XX  
CC The invention relates to a method for modulating activity of a thyroid  
CC hormone receptor that comprises administration of an aromatic compound  
CC which fits spatially and preferentially into a thyroid hormone ligand

CC binding domain. The aromatic compound (of a specified formula) can be  
CC used to increase alpha-glucophosphate dehydrogenase (GPDH) levels, at  
CC levels which do not significantly modify cardiac GPDH levels and are  
CC indicated in the treatment of obesity. The compound also lower total  
CC plasma cholesterol and triglyceride levels and can be used as anti-  
CC hypertriglyceridaemic agents. The compound may also be used for treating  
CC atherosclerosis and may be indicated in thyroid hormone replacement  
CC therapy in patients with compromised cardiac function. Sequences  
CC AAY21621-636 amino acid sequences of ligand binding domains of several  
CC members of the nuclear receptor superfamily.  
XX  
SQ Sequence 468 AA;

Query Match 68.9%; Score 31; DB 20; Length 468;  
Best Local Similarity 60.0%; Pred. NO: 1.5e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 31 mgniqueisqs 40  
||:| :|||

RESULT 11  
AAY05470  
ID AAY05470 standard; Protein: 468 AA.  
XX  
AC AAY05470;  
XX  
DT 07-JUL-1999 (first entry)  
XX  
DE Human PPAR-alpha protein sequence.  
XX  
KW Nuclear receptor agonist; antagonist; identification; PPAR;  
KW peroxisome proliferator activated receptor.  
XX  
OS Homo sapiens.  
XX WO9918124-A1.  
PN  
XX  
PD 15-APR-1999.  
XX  
PF 06-OCT-1998; 98WO-US21049.  
XX  
PR 07-OCT-1997; 97US-0061385.  
XX  
PA (MERI ) MERCK & CO INC.  
XX  
PI Cummings RT, Hermes JD, Moller DE, Zhou G;  
XX WPI; 1999-263998/22.  
DR N-PSDB; AAX36521.  
XX  
XX Identifying nuclear receptor agonists and antagonists  
XX Disclosure; Fig 8a; 60pp; English.  
PS  
CC This sequence is the human peroxisome proliferator activated  
CC receptor-alpha (PPAR-alpha).  
CC The invention relates to a method for identifying nuclear receptor  
CC agonists and antagonists comprises measuring fluorescent resonance energy  
CC transfer between fluorescent-labelled nuclear receptors and  
CC co-activators. The method can be used for identifying agonists and  
CC antagonist of nuclear receptors.  
XX  
SQ Sequence 468 AA;

Query Match 68.9%; Score 31; DB 20; Length 468;  
Best Local Similarity 60.0%; Pred. NO: 1.5e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10

Db 31 mgniqueisqs 40  
||:| :|||

RESULT 12  
AAB20342  
ID AAB20342 standard; Protein: 468 AA.  
XX  
AC AAB20342;  
XX  
DT 29-MAY-2001 (first entry)  
XX  
DE Peroxisome proliferator-activated receptor alpha.  
XX  
KW Peroxisome proliferator-activated receptor alpha; PPAR-alpha;  
KW human; steroid hormone receptor; lipid metabolism;  
KW low density lipoprotein; cholesterol; apolipoprotein;  
KW atherosclerosis; hyperapobetalipoproteinemia;  
KW coronary heart disease; diagnosis; prognosis; therapy.  
XX  
OS Homo sapiens.  
XX WO200120037-A2.  
PN  
XX  
PD 22-MAR-2001.  
XX  
PF 14-SEP-2000; 2000WO-US25124.  
XX  
PR 17-SEP-1999; 99US-0154736.  
XX  
PA (COMP-) COMPLEXE HOSPITALIER SAGAMIE.  
PA (UYMC-) UNIV MCGILL.  
XX  
PI Hudson TJ, Gaudet D, Vohl M, Brewer C, Morgan K;  
XX N-PSDB; AAF30517.  
DR  
XX  
PT Prognosis of a disorder of lipid metabolism in an individual involves  
PT detecting single nucleotide polymorphisms in codon 162 and 253 of  
PT peroxisome proliferator-activated receptor-alpha gene  
XX  
PS Claim 21; Fig 3; 52pp; English.  
XX  
CC The present sequence is that of human peroxisome  
CC proliferator-activated receptor alpha (PPAR-alpha), a member of  
CC the steroid hormone receptor superfamily that is involved in the  
CC control of cellular lipid utilization. A L162V mutation has been  
CC identified in the protein, which results from a missense mutation  
CC in exon 5 of the PPAR-alpha gene (see AAF30517). The mutation is  
CC within the conserved DNA binding or C domain of the protein. The  
CC presence of the rare L162 allele is associated with a greater  
CC likelihood of an individual having, or of developing, a disorder  
CC of lipid metabolism, especially high low density lipoproteinemia  
CC cholesterol levels, atherosclerosis, hyperapobetalipoproteinemia  
CC and/or coronary heart disease. PPAR-alpha polypeptides and  
CC polynucleotides are used in claimed methods for predicting the  
CC likelihood that an individual will have a disorder of lipid  
CC metabolism, and for treating such a disorder (by administering a  
CC functional portion of PPAR-alpha or a nucleic acid encoding it).  
XX  
SQ Sequence 468 AA;

Query Match 68.9%; Score 31; DB 22; Length 468;  
Best Local Similarity 60.0%; Pred. NO: 1.5e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 31 mgniqueisqs 40  
||:| :|||



RESULT 13  
AAG11352  
ID AAG11352 standard; Protein; 288 AA.  
XX  
AC AAG11352;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10036.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match. Score 30; DB 21; Length 288;  
Best Local Similarity 77.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GSIQVLSQS 10  
Db 142 ggiivalsqs 150  
RESULT 14  
AAG16074  
ID AAG16074 standard; Protein; 288 AA.  
XX  
AC AAG16074;  
XX  
DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 16574.  
DE  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 30; DB 21; Length 288;  
Best Local Similarity 77.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GSIVQLSQS 10  
| || ||||  
Db 142 ggivalsqs 150

RESULT 15  
AAG11351  
ID AAG11351 standard; Protein; 303 AA.  
XX  
AC AAG11351;

DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10035.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS	Arabidopsis thaliana.		
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	99US-0141287.
PR	05-MAR-1999;	99US-0123180.	99US-0141842.
PR	09-MAR-1999;	99US-0123548.	99US-0142154.
PR	23-MAR-1999;	99US-0125788.	99US-0142055.
PR	25-MAR-1999;	99US-0126264.	99US-0142390.
PR	29-MAR-1999;	99US-0126785.	99US-0142803.
PR	01-APR-1999;	99US-0127462.	99US-0142920.
PR	06-APR-1999;	99US-0128234.	99US-0142977.
PR	08-APR-1999;	99US-0128714.	99US-0143542.
PR	16-APR-1999;	99US-0129845.	99US-0143624.
PR	19-APR-1999;	99US-0130077.	99US-0144005.
PR	21-APR-1999;	99US-0130449.	99US-0144085.
PR	23-APR-1999;	99US-0130510.	99US-0144086.
PR	28-APR-1999;	99US-0130891.	99US-0144325.
PR	30-APR-1999;	99US-0131449.	99US-0144331.
PR	30-APR-1999;	99US-0132048.	99US-0144332.
PR	30-APR-1999;	99US-0132407.	99US-0144333.
PR	04-MAY-1999;	99US-0132484.	99US-0144334.
PR	05-MAY-1999;	99US-0132485.	99US-0144335.
PR	06-MAY-1999;	99US-0132486.	99US-0144352.
PR	07-MAY-1999;	99US-0132487.	99US-0144632.
PR	11-MAY-1999;	99US-0132863.	99US-0144884.
PR	14-MAY-1999;	99US-0134256.	99US-0144814.
PR	14-MAY-1999;	99US-0134218.	99US-0145086.
PR	14-MAY-1999;	99US-0134219.	99US-0145088.
PR	14-MAY-1999;	99US-0134221.	99US-0145085.
PR	18-MAY-1999;	99US-0134370.	99US-0145087.
PR	19-MAY-1999;	99US-0134768.	99US-0145089.
PR	20-MAY-1999;	99US-0134941.	99US-0145192.
PR	21-MAY-1999;	99US-0135353.	99US-0145145.
PR	24-MAY-1999;	99US-0135629.	99US-0145218.
PR	25-MAY-1999;	99US-0136021.	99US-0145224.
PR	27-MAY-1999;	99US-0136392.	99US-0145276.
PR	28-MAY-1999;	99US-0136782.	99US-0145913.
PR	01-JUN-1999;	99US-0137222.	99US-0145918.
PR	03-JUN-1999;	99US-0137528.	99US-0145919.
PR	04-JUN-1999;	99US-0137502.	99US-0146386.
PR	07-JUN-1999;	99US-0137724.	99US-0146388.
PR	08-JUN-1999;	99US-0138094.	99US-0146389.
PR	10-JUN-1999;	99US-0138540.	99US-0147038.
PR	14-JUN-1999;	99US-0139119.	99US-0147204.
PR	16-JUN-1999;	99US-0139452.	99US-0147302.
PR	16-JUN-1999;	99US-0139453.	99US-0147192.
PR	17-JUN-1999;	99US-0139492.	99US-0147260.
PR	18-JUN-1999;	99US-0139454.	99US-0147303.
PR	18-JUN-1999;	99US-0139455.	99US-0147416.
PR	18-JUN-1999;	99US-0139456.	99US-0147493.
PR	18-JUN-1999;	99US-0139457.	99US-0147935.
PR	18-JUN-1999;	99US-0139458.	99US-0148171.
PR	18-JUN-1999;	99US-0139459.	99US-0148319.
PR	18-JUN-1999;	99US-0139460.	99US-0148341.
PR	18-JUN-1999;	99US-0139461.	99US-0148565.
PR	18-JUN-1999;	99US-0139462.	99US-0148684.
PR	18-JUN-1999;	99US-0139463.	99US-0149368.
PR	18-JUN-1999;	99US-0139750.	99US-0149175.
PR	21-JUN-1999;	99US-0139763.	99US-0149426.
PR	22-JUN-1999;	99US-0139817.	99US-0149722.
PR	23-JUN-1999;	99US-0139899.	99US-0149723.
PR	23-JUN-1999;	99US-0140353.	99US-0149929.
PR	24-JUN-1999;	99US-0140354.	99US-0149902.
PR	28-JUN-1999;	99US-0140695.	99US-0149930.
PR	29-JUN-1999;	99US-0140823.	99US-0149902.
PR			99US-0150566.
			99US-0150884.
			99US-0151065.
			99US-0151066.
			99US-0151080.
			99US-0151303.
			99US-0151438.
			99US-0151930.
			99US-0152363.
			99US-0153070.
			99US-0153758.



PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 30; DB 21; Length 303;  
Best Local Similarity 77.8%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
I || ||||  
Db 157 ggivalsqs 165

RESULT 16  
AAG16073  
ID AAG16073 standard; Protein; 303 AA.

XX AC AAG16073;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 16573.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.



PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.





PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 23-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 30; DB 21; Length 425;  
Best Local Similarity 77.8%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
| | | | |  
Db 279 ggivalsqs 287

RESULT 19  
AAB26599  
ID AAB26599 standard; protein; 750 AA.  
XX  
AC AAB26599;  
XX  
DT 01-FEB-2001 (first entry)  
XX  
DE Synechocystis sp phytochrome-related gene Cph8.  
XX  
KW Phytochrome; fluorescent apophytochrome; bilin; Cph.  
XX  
OS Synechocystis sp.  
XX  
PN WO200056355-A1.  
XX  
PD 28-SEP-2000.  
XX  
PF 14-MAR-2000; 2000WO-US06607.  
XX  
PR 19-MAR-1999; 99US-0272809.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Lagarias JC;  
XX  
DR WPI; 2000-602195/57.  
XX  
PT Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from  
PT Synechocystis species) and phycoerythrobilin conjugate, useful as  
PT fluorescent markers for biological research -  
XX  
PS Example 1; Page 50; 52pp; English.  
XX

CC The present invention is related to fluorescent apophytochrome-bilin  
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from  
CC Synechocystis species is used as the apophytochrome and the bilin is  
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent  
CC markers for biological research. The phytofluors have a long wavelength  
CC absorption maxima, a high molar absorption coefficient and the  
CC recombinant apoproteins can spontaneously assemble with a variety of  
CC bilin chromophore precursors. The present sequence is a phytochrome

CC related protein from Synechocystis sp.

XX Sequence 750 AA;

Best Local Similarity 66.7%; Score 30; DB 21; Length 750; 0; Gaps 0;  
Matches 6; Conservative 2; Mismatches 1; Indels 1; Indels 2; Indels 0;

Query Match 66.7%; Score 30; DB 21; Length 750;  
Best Local Similarity 66.7%; Pred. No. 4.2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 1; Indels 2; Indels 0;

OY 1 MGSIVQLSQ 9  
: ||:||||  
Db 437 inslqlsq 445

RESULT 20  
AAAY41012  
ID AAY41012 standard; Protein; 2559 AA.  
XX AC  
AC AAY41012;

XX 06-DEC-1999 (first entry)

DE Amino acid sequence of DRHOGEF2.

XX RGS; guanine nucleotide exchange factor; GEF; RGS domain; GAS; cancer;  
KW G protein alpha subunit; cell proliferation; growth control; hemostasis;  
KW morphogenesis; stress fiber formation; integrin-mediated interaction;  
KW embryonic development; tumor cell growth; leukocyte homing;  
KW bone resorption; clot retraction; db1 homology domain; mechanical stress;  
KW pleckstrin homology domain; DRHOGEF2.

XX Drosophila sp.

XX WO9947557-A2.

XX 23-SEP-1999.

PF 18-MAR-1999; 99WO-US06051.

PR 18-MAR-1998; 98US-0078634.

PA (ONYX-) ONYX PHARM INC.

PI Bollag G, Hart MJ, Roscoe W, Polakis P, Sternweis P, Kozasa T;  
PI Jiang X;

DR WPI; 1999-571821/48.

DR N-PSDB; AAZ22284.

XX New isolated RGS-GEF polypeptides, used to develop products for

modulating, e.g. cell proliferation and integrin-mediated interactions

XX Examples; Fig 16; 75pp; English.

XX The invention relates to isolated RGS-guanine nucleotide exchange factor  
(GEF) polypeptides. The novel isolated RGS-GEF polypeptide consists of an  
RGS domain of a GEF protein and does not comprise a db1 homology (DH)  
domain or a pleckstrin homology (PH) domain. The RGS-GEF polypeptides can  
be used for modulating an activity of a G protein alpha subunit (GAS).  
The products can be used for the regulation of biological pathways in  
which a RGS-GEF polypeptide is involved, particularly pathological  
conditions, e.g. cell proliferation (e.g. cancer), morphogenesis, growth  
control, stress fiber formation, and integrin-mediated interactions, such  
as embryonic development, tumor cell growth and metastasis, programmed  
cell death, hemostasis, leukocyte homing and activation, bone resorption,  
clot retraction, and the response of cells to mechanical stress. The  
products can also be used for detection, diagnosis and production of  
transgenic animals. The present sequence represents the amino acid  
sequence of DRHOGEF2.

XX Sequence 2559 AA;

Query Match 66.7%; Score 30; DB 20; Length 2559;  
Best Local Similarity 70.0%; Pred. No. 1.7e+03;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10  
: ||:||||  
Db 196 mgggvqlnqs 205

RESULT 21

AAW52846  
ID AAW52846 standard; Protein; 5069 AA.  
XX AC  
AC AAW52846;

XX 24-JUL-1998 (first entry)

XX A. mediterranei rifamycin synthesis gene cluster fragment protein B.  
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;  
KW polyketide synthase; actinomycete; ansamycin.

XX Amycolatopsis mediterranei.

XX WO9807868-A1.

XX 26-FEB-1998.

PF 18-AUG-1997; 97WO-EP04495.

PR 20-AUG-1996; 96EP-0810551.

XX (NOVS) NOVARTIS AG.

XX Engel N, Schupp T, Toupet C;

DR WPI; 1998-169172/15.

DR N-PSDB; AAV21187.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used

to produce rifamycin and rifamycin analogues

XX Claim 6; Page 126-151; 205pp; English.

XX The present sequence represents a Amycolatopsis mediterranei rifamycin  
synthesis gene cluster ORF B protein from the present invention. The  
DNA fragment comprises a DNA region involved directly or indirectly  
in the gene cluster responsible for rifamycin synthesis, including  
the adjacent DNA regions to the right and left which, by reason of  
their function in connection with rifamycin biosynthesis, qualify  
as constituents of this rifamycin gene cluster, and functional  
fragments, derivatives or constituents of these. The Amycolatopsis  
mediterranei rifamycin synthesis gene cluster DNA fragment can be used  
for producing rifamycin, rifamycin analogues or precursors. It can also  
be used for inactivating or modifying genes involved in ansamycin or  
rifamycin biosynthesis. The DNA can be used for constructing mutant  
actinomycetes strains from which the natural rifamycin or ansamycin  
biosynthesis gene cluster has been partly or completely deleted. The  
DNA fragment can be used for assembling a library of polyketide  
synthases, which can be used for assembling a library of polyketides.  
A hybridisation probe of the invention can be used for identifying DNA  
fragments involved in the biosynthesis of ansamycins.

XX Sequence 5069 AA;

Query Match 66.7%; Score 30; DB 19; Length 5069;  
Best Local Similarity 50.0%; Pred. No. 3.6e+03;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10  
: ||:||||

Db 4219 lgavvelpqs 4228

RESULT 22  
AAB65054

ID AAB65054 standard; peptide; 36 AA.  
XX  
AC AAB65054;  
XX  
DT 23-MAR-2001 (first entry)  
XX  
DE Gene #7 associated peptide #8.  
XX  
KW Secreted protein; gene therapy; vaccine; cancer; leukemia;  
KW autoimmune disease; allergy; inflammation; graft rejection;  
KW hyperproliferation; cardiovascular; infection.  
XX  
OS Homo sapiens.  
XX  
PN WO200075375-A1.  
XX  
PD 14-DEC-2000.  
XX  
PF 02-JUN-2000; 2000WO-US15187.  
XX  
PR 07-JUN-1999; 99US-0137725.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;  
PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Ni J, Young PE;  
XX  
DR WPI; 2001-061741/07.  
XX  
XX Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for  
PT preventing, diagnosing and/or treating cancers and for promoting wound  
PT healing -  
XX  
PS Disclosure; Page 29; 530pp; English.  
XX  
CC The present invention relates to 26 secreted human proteins. The  
CC proteins may be used in the prevention, diagnosis and treatment of  
CC diseases associated with inappropriate polypeptide expression.  
CC For example, they may be used in gene therapy or in vaccines.  
CC Typical of diseases which are potentially treatable are cancers  
CC (including leukemia), autoimmune diseases, allergies, inflammation,  
CC graft rejection, hyperproliferation, cardiovascular diseases  
CC (particularly critical limb ischemia and coronary disease) and any  
CC involving abnormal angiogenesis, neurodegeneration and/or  
XX infectious diseases.  
XX  
SQ Sequence 36 AA;

Query Match 64.4%; Score 29; DB 22; Length 36;  
Best Local Similarity 55.6%; Pred. No. 23;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSVQLSQS 10  
II:::III:  
Db 9 gsvrlsq 17

RESULT 23  
AAW55323

ID AAW55323 standard; Protein; 155 AA.  
XX  
AC AAW55323;  
XX  
DT 15-JUN-1998 (first entry)  
XX  
DE H. pylori ORF hp2el185orf6 protein.  
XX

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
XX inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
OS Helicobacter pylori.  
XX  
PN WO9737044-A1.  
XX  
PD 09-OCT-1997.  
XX  
PF 27-MAR-1997; 97WO-US05223.  
XX  
PR 06-DEC-1996; 96US-0761318.  
PR 29-MAR-1996; 96US-0625811.  
PR 02-APR-1996; 96US-0758731.  
PR 25-OCT-1996; 96US-0736905.  
PR 28-OCT-1996; 96US-0738859.  
XX  
PA (ASTR ) ASTRA AB.  
XX  
PI Alm RA, Smith D;  
XX  
DR WPI; 1997-503122/46.  
DR N-PSDB; AAV24732.  
XX  
XX Helicobacter pylori nucleic acid sequences and encoded  
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection  
XX  
PS Claim 14; Page 550; 1145pp; English.  
XX  
CC This sequence is a H. pylori protein of unspecified function.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors. The  
CC DNA and probes derived from it may be used for the identification of  
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
CC acid sequences complementary to the DNA act as antisense sequences and  
CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
CC against the protein can be used in immunoassays to evaluate the abundance  
CC and distribution of H. pylori-specific antigens. The genomic sequence of  
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
CC by mechanically shearing the bacterial DNA. The sequences were analysed  
CC for ORF of at least 180 nucleotides, and the predicted coding regions  
CC defined by computer evaluation. To identify likely H. pylori antigens for  
CC vaccine development, the amino acid sequences predicted from various ORF  
CC were analysed for significant homology to other known or exported  
CC membrane proteins. Having identified and determined the sequences of  
CC interest, particular regions can be isolated from H. pylori by PCR  
CC amplification for recombinant polypeptide production, e.g. in E. coli  
XX hosts.  
XX  
SQ Sequence 155 AA;

Query Match 64.4%; Score 29; DB 18; Length 155;  
Best Local Similarity 55.6%; Pred. No. 1.2e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:II::III:  
Db 101 igsvvldq 109

RESULT 24  
AAY60166

ID AAY60166 standard; Protein; 160 AA.  
XX  
AC AAY60166;  
XX  
DT 31-JAN-2000 (first entry)  
XX  
DE Human endometrium tumour EST encoded protein 226.

XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
KW treatment; uterine; gene therapy; expressed sequence tag.  
XX Homo sapiens.  
XX DE19817948-A1.  
XX 21-OCT-1999.  
XX 17-APR-1998; 98DE-1017948.  
XX 17-APR-1998; 98DE-1017948.  
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
PI WPI; 1999-591957/51.  
XX N-PSDB; AAZ42055.  
XX New nucleic acid sequences expressed in uterine cancer tissues, and  
PT derived polypeptides, for treatment of uterine and endometrial cancer  
PT and identification of therapeutic agents -  
XX Claim 23; Page 365; 444pp; German.  
XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for  
CC generation of specific antibodies. (A) are identified by assembling ESTs  
CC (expressed sequence tags) from a particular tissue type before comparison  
CC of expression patterns. This allows a significantly longer fragment of  
CC the gene to be revealed, so should reduce the number of failures  
CC associated with the fact that ESTs from different libraries may represent  
CC different parts of the same unknown gene, distorting the estimated  
CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent  
CC protein fragments encoded by the human endometrium tumour cDNA library  
CC derived EST fragments represented in AAZ41981-242121.  
XX Sequence 160 AA;  
SQ  
Query Match 64.4%; Score 29; DB 20; Length 160;  
Best Local Similarity 55.6%; Pred. NO. 1.2e+02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GSIVQLSQS 10  
DB 42 gsvlrslsq 50  
RESULT 25  
ID AAB42674 standard; Protein; 230 AA.  
XX AAB42674;  
XX 08-FEB-2001 (first entry)  
XX Human ORFX ORF2438 polypeptide sequence SEQ ID NO:4876.  
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoiatric; antiparkinsonian; neurotropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX Homo sapiens.  
XX OS  
XX WO200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US08621.  
XX 31-MAR-1999; 99US-0127607.  
XX 02-APR-1999; 99US-0127636.  
XX 05-APR-1999; 99US-0127728.  
XX 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX N-PSDB; AAC76883.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 11; Page 4056-4057; 5507pp; English.  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoiatric; antiparkinsonian; neurotropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, diabetes mellitus,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX Sequence 230 AA;  
SQ  
Query Match 64.4%; Score 29; DB 21; Length 230;  
Best Local Similarity 55.6%; Pred. NO. 1.8e+02;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GSIVQLSQS 10  
DB 112 gsvlrslsq 120  
RESULT 26  
AAR85494  
ID AAR85494 standard; Protein; 240 AA.  
XX AAR85494;  
AC



XX 16-MAR-1996 (first entry)  
XX ErB-2-specific scFv(FRP)5.  
DE Single chain antibody; scFv; antibody engineering; antitumour;  
XX tumour antigen binding; T-cell receptor; cytotoxic T-lymphocyte;  
KW cell targeting; monoclonal antibody; erbB-2; cancer;  
KW adoptive immunotherapy.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Region 2..30  
FT /label= FR1  
FT Region 31..35  
FT /label= CDR1H  
FT Region 36..49  
FT /label= FR2  
FT Region 50..66  
FT /label= CDR2H  
FT Region 67..98  
FT /label= FR3  
FT Region 99..108  
FT /label= CDR3H  
FT Region 110..120  
FT /label= FR4  
FT Peptide 121..134  
FT /label= Spacer\_peptide  
FT Region 135..157  
FT /label= FR6  
FT Region 158..168  
FT /label= CDR1L  
FT Region 169..183  
FT /label= FR7  
FT Region 184..190  
FT /label= CDR3L  
FT Region 191..222  
FT /label= FR8  
FT Region 223..231  
FT /label= CDR3L  
FT Region 232..240  
FT /label= FR9  
XX WO9530014-A1.  
PN  
XX  
PD 09-NOV-1995.  
XX  
XX 20-APR-1995; 95WO-EP01494.  
PF  
XX 02-MAY-1994; 94EP-0810244.  
PR  
XX (CIBA ) CIBA GEIGY AG.  
PA  
XX Groner B, Moritz D;  
PI  
XX WPI; 1995-393085/50.  
DR N-PSDB; AAT05781.  
DR  
XX New bifunctional proteins for use in killing tumour cells - contg. a  
PT tumour antigen binding domain, a hinge region and a zeta chain  
PT derived from a T-cell antigen receptor  
PT  
XX Disclosure; Page 23-25; 46pp; English.  
PS  
XX Single chain antibody scFv(FRP)5 (AAR85494) is specific for the  
CC extracellular domain of erbB-2. The heavy and light chain variable  
CC regions of the scFv are derived from mouse monoclonal antibody FRP5  
CC (ECACC 90112115), while the framework regions are of human origin.  
CC The scFv is used as the tumour antigen binding domain of a  
CC bifunctional protein also contg. a hinge region and a zeta chain of a  
CC T-cell receptor that is expressed in host cells, esp. cytotoxic  
CC T-lymphocytes (CTLs). The CTLs are provided with a defined tumour

CC cell specificity enabling targeting and destruction of the tumour  
CC cells.  
XX  
SQ Sequence 240 AA;  
Query Match 64.4%; Score 29; DB 16; Length 240;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GSIVQLSQS 10  
Db 133 gsdqltqs 141  
RESULT 27  
AAR85495  
ID AAR85495 standard; Protein; 240 AA.  
XX  
AC AAR85495;  
XX 16-MAR-1996 (first entry)  
DT  
DE ScFv(FWP51).  
XX  
KW Single chain antibody; scFv; antibody engineering; antitumour;  
KW tumour antigen binding; T-cell receptor; cytotoxic T-lymphocyte;  
KW cell targeting; monoclonal antibody; cancer;  
KW adoptive immunotherapy.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Region 2..30  
FT /label= FR1  
FT Region 31..35  
FT /label= CDR1H  
FT Region 36..49  
FT /label= FR2  
FT Region 50..66  
FT /label= CDR2H  
FT Region 67..98  
FT /label= FR3  
FT Region 99..109  
FT /label= CDR3H  
FT Region 110..120  
FT /label= FR4  
FT Peptide 121..135  
FT /label= Spacer\_peptide  
FT Region 136..158  
FT /label= FR6  
FT Region 159..164  
FT /label= CDR1L  
FT Region 170..184  
FT /label= FR7  
FT Region 185..191  
FT /label= CDR3L  
FT Region 192..223  
FT /label= FR8  
FT Region 224..231  
FT /label= CDR3L  
FT Region 232..240  
FT /label= FR9  
XX WO9530014-A1.  
PN  
XX 09-NOV-1995.  
PD  
XX 20-APR-1995; 95WO-EP01494.  
PF  
XX 02-MAY-1994; 94EP-0810244.  
PR  
XX (CIBA ) CIBA GEIGY AG.  
PA

XX PI Groner B, Moritz D;  
XX WPI; 1995-393085/50.  
DR N-PSDB; AAT05782.  
XX  
PT New bifunctional proteins for use in killing tumour cells - contg. a  
PT tumour antigen binding domain, a hinge region and a zeta chain  
PT derived from a T-cell antigen receptor  
XX  
PS Disclosure; page 26-27; 46pp; English.  
XX  
CC Single chain antibody scFv(FWP51) (AAR85495) has heavy and light chain  
CC variable regions derived from mouse monoclonal antibody FRP5  
CC (ECACC 90112115) and framework regions of human origin. The scFv is  
CC used as the tumour antigen binding domain of a bifunctional protein  
CC also contg. a hinge region and a zeta chain of a T-cell receptor that  
CC is expressed in host cells, esp. cytotoxic T-lymphocytes (CTLs). The  
CC CTLs are provided with a defined tumour cell specificity enabling  
CC targeting and destruction of the tumour cells.  
XX  
SQ Sequence 240 AA;  
Query Match 64.4%; Score 29; DB 16; Length 240;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GSIVQLSOS 10  
Db 134 gsdigtqts 142  
RESULT 28  
AAW05133  
ID AAW05133 standard; Protein; 240 AA.  
XX AC AAW05133;  
XX DT 29-JAN-1997 (first entry)  
XX DE Single chain antibody scFv(225).  
XX KW Single chain antibody; scFv; monoclonal antibody; MAb; EGF;  
XX KW epidermal growth factor; receptor; antitumour; cancer; therapy.  
XX OS Chimeric Mus sp.;  
XX OS Chimeric synthetic.  
XX FH Key Location/Qualifiers  
FT Region 1..119  
FT /label= VH\_region  
FT /note= "monoclonal antibody 225 VH"  
FT Peptide 120..133  
FT /label= Linker  
FT /note= "synthetic spacer peptide"  
FT Region 134..240  
FT /label= VL  
FT /note= "monoclonal antibody 225 VL"  
XX PN EP739984-A1.  
XX 30-OCT-1996.  
XX PD 26-APR-1995; 95EP-0106275.  
XX PF 26-APR-1995; 95EP-0106275.  
XX PR 26-APR-1995; 95EP-0106275.  
XX PA (SANT-) SAN TUMORFORSCHUNGS GMBH.  
XX PI Groner B, Schmidt M, Wels W;  
XX WPI; 1996-478748/48.  
DR

XX N-PSDB; AAT42033.  
XX Bivalent fusion proteins that bind epidermal growth factor receptor  
PT or analogues - and comprise at least two different cell surface  
PT binding domain(s), useful for tumour therapy  
XX  
PS Example 2; Page 17; 52pp; English.  
XX  
CC scFv(225) (AAW05133) comprises the single-chain binding region of  
CC murine monoclonal antibody 225, which is specific for the human  
CC epidermal growth factor receptor. It is encoded by plasmid  
CC pMW152-225 (see also AAT42033), constructed by cloning MAb 225  
CC VH and VL region cDNAs into plasmid pMW152. Novel bivalent  
CC proteins (see also AAW05134-44), some of them including scFv(225)  
CC and an effector e.g. cytotoxin, can be produced in bacterial  
CC host cells, and are useful as antitumour agents.  
XX  
SQ Sequence 240 AA;  
Query Match 64.4%; Score 29; DB 17; Length 240;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GSIVQLSOS 10  
Db 133 gsdigtqts 141  
RESULT 29  
AAW05134  
ID AAW05134 standard; Protein; 240 AA.  
XX AC AAW05134;  
XX DT 29-JAN-1997 (first entry)  
XX DE Single chain antibody scFv(FRP5).  
XX KW Single chain antibody; scFv; monoclonal antibody; MAb; EGF; erbB-2;  
XX KW epidermal growth factor; receptor; antitumour; cancer; therapy.  
XX OS Chimeric Mus sp.;  
XX OS Chimeric synthetic.  
XX FH Key Location/Qualifiers  
FT Region 1..119  
FT /label= VH\_region  
FT /note= "monoclonal antibody FRP5 VH"  
FT Peptide 120..133  
FT /label= Linker  
FT /note= "synthetic spacer peptide"  
FT Region 134..240  
FT /label= VL  
FT /note= "monoclonal antibody FRP5 VL"  
XX PN EP739984-A1.  
XX 30-OCT-1996.  
XX PD 26-APR-1995; 95EP-0106275.  
XX PF 26-APR-1995; 95EP-0106275.  
XX PR 26-APR-1995; 95EP-0106275.  
XX PA (SANT-) SAN TUMORFORSCHUNGS GMBH.  
XX PI Groner B, Schmidt M, Wels W;  
XX WPI; 1996-478748/48.  
DR N-PSDB; AAT42034.  
XX  
PT Bivalent fusion proteins that bind epidermal growth factor receptor  
PT or analogues - and comprise at least two different cell surface

PT binding domain(s), useful for tumour therapy

XX

PS Example 3; Page 18; 52pp; English.

XX

CC scFv(FRP5) (AAW05134) comprises the single-chain binding region of

CC murine monoclonal antibody FRP5, which is specific for the human

CC epidermal growth factor receptor erbB-2 (Neu/HER-2). It is encoded

CC by plasmid pW152-5 (see also AAT42034). Novel bivalent proteins

CC (see also AAW05135-44), some of them including scFv(FRP5) and an

CC effector, e.g. cytotoxin, molecule can be produced in bacterial

CC host cells, and are useful as antitumour agents.

XX

SQ Sequence 240 AA;

Query Match 64.4%; Score 29; DB 17; Length 240;

Best Local Similarity 66.7%; Pred. No. 1.9e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

DB 133 gsdiglqtgs 141

II :||:|

RESULT 30

AAW25784

ID AAW25784 standard; Protein; 240 AA.

XX

AC AAW25784;

XX

DT 22-DEC-1997 (first entry)

XX

DE Anti-gp130 antibody derived scFv.

XX

KW Single chain Fv; scFv; antibody; E. coli; anti-T3 antibody;

KW anti-gp130 antibody; inclusion body; chaperonin.

XX

OS Synthetic.

XX

PN JP09220092-A.

XX

PD 26-AUG-1997.

XX

PF 15-FEB-1996; 96JP-0027622.

XX

PR 15-FEB-1996; 96JP-0027622.

XX

PA (TOYJ ) TOSOH CORP.

XX

DR WPI; 1997-474306/44.

DR N-PSDB; AAT91615.

XX

PT Producing single chain Fv antibody in Escherichia coli - by

PT expression in an inclusion body, followed by protein folding or by

PT co-expression with a chaperonin as a soluble fraction

XX

PS Claim 4; Page 6-7; 9pp; Japanese.

XX

CC The sequences given in AAW25783-84 represent single chain Fv (scFv)

CC antibodies which are produced in E.coli. The scFV's are derived

CC from an anti-T3 antibody or an anti-gp130 antibody and are produced

CC either by: (1) expression as an inclusion body, followed by folding

CC (i.e. by denaturation and solubilisation) or (2) expression as a

CC soluble fraction by co-expression with chaperonin. Using the methods,

CC scFv can be produced in E. coli, either in a soluble fraction or in

CC inclusion bodies. Like the parent double chain antibodies, the scFv

CC polypeptides are useful in immunological diagnosis or for separation

CC and purification, but they lessen the side effects caused by antibody

CC constant regions. The availability of scFv polypeptides contributes

CC to fundamental research and development of diagnostic and therapeutic

CC drugs.

XX

SQ Sequence 240 AA;

Query Match 64.4%; Score 29; DB 18; Length 240;

Best Local Similarity 77.8%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

DB 119 gsdvqlqqs 127

II :||:|

RESULT 31

AAW71243

ID AAW71243 standard; Protein; 240 AA.

XX

AC AAW71243;

XX

DT 20-NOV-1998 (first entry)

XX

DE scFv comprising heavy and light chain variable domains of antibody 14E1.

XX

KW Antibody 14E1; human epidermal growth factor receptor; EGF-R; treatment;

KW carcinoma; EGF-RvIII; in vitro purging; breast carcinoma.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Protein 1..119

FT /note= "Heavy chain variable domain"

FT Peptide 120..134

FT /note= "synthetic linker"

FT Protein 135..240

FT /note= "light chain variable domain"

XX

PN WO9836074-A2.

XX

PD 20-AUG-1998.

XX

PF 13-FEB-1998; 98WO-IB00413.

XX

PR 13-FEB-1997; 97US-0800198.

XX

PA (SCHD ) SCHERING AG.

XX

PI Schmidt M, Schneider DW, Vakalopoulou E, Wells WS;

XX

DR WPI; 1998-467176/40.

DR N-PSDB; AAV54790.

XX

PT New single chain polypeptide that binds to epidermal growth factor

PT receptor - derived from antibody 14E1, and related nucleic acid and

PT double-chain forms, used for treatment and diagnosis of cancer or

PT other cell proliferative diseases

XX

PS Claim 11; Fig 12; 48pp; English.

XX

CC The present sequence represents a single-chain polypeptide (scFv) of

CC the invention. It is derived from murine monoclonal antibody 14E1. This

CC antibody is active against the human epidermal growth factor receptor

CC (EGF-R). The heavy and light chains of antibody 14E1 are used to

CC construct a scFv that has binding affinity for an EGF-R. The scFv

CC comprises two segments which contain the binding parts of the heavy and

CC light chain variable domains of the monoclonal antibody 14E1. The scFvs

CC of the invention are used to treat carcinomas (gliomas, melanomas and

CC tumours) that overexpress EGF-R, specifically EGF-RvIII, relative to

CC normal cells, particularly by blocking proliferation of such cells by

CC preventing receptor activation. They can also be used for in vitro

CC purging. When the scFv is appropriately labelled, it can be used

CC diagnostically to image cells that express EGF-R, specifically breast

CC carcinoma.

XX

SQ Sequence 240 AA;

Query Match 64.4%; Score 29; DB 19; Length 240;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
Db 133 gsdqltqs 141

RESULT 32  
AAR26979  
ID AAR26979 standard; Protein; 241 AA.  
XX AAR26979;  
XX  
DT 11-FEB-1993 (first entry)  
XX  
DE Fv heavy chain/light chain variable domain fusion protein.  
XX  
KW Minigene; pM218/1; monoclonal antibody; kappa; light chain;  
KW variable region; pMW31; tumour.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT Domain 2..120 /label= FRP5\_heavy\_chain\_variable\_domain  
FT Region 32..36 /label= CDR1H  
FT Region 51..67 /label= CDR2H  
FT Region 100..109 /label= CDR3H  
FT Peptide 121..135 /label= Linker  
FT Domain 136..241 /label= FRP5\_light\_chain\_variable\_domain  
FT Region 159..169 /label= CDR1L  
FT Region 185..191 /label= CDR2L  
FT Region 224..232 /label= CDR3L  
XX EP502812-A.  
PN  
XX  
XX 09-SEP-1992.  
XX  
XX 27-JAN-1992; 92EP-0810056.  
XX  
XX 05-FEB-1991; 91EP-0810079.  
XX  
XX (CIBA ) CIBA GEIGY AG.  
XX  
XX Groner B, Hardman N, Harwerth I, Hynes NE, Wells WS;  
XX Zwickl M;  
XX WPI; 1992-302096/37.  
XX N-PSDB; AAQ28256.  
XX  
XX Recombinant antibodies directed to growth factor receptor C-erbB-2 -  
XX for diagnosing and treating tumours expressing C-erbB-2 e.g. breast  
XX or ovarian tumours  
XX  
XX Disclosure; Page 31-33; 67pp; English.  
XX  
XX The sequence given is the Fv heavy chain/light chain variable domain  
XX fusion protein which binds to the extracellular domain of the growth  
XX factor receptor c-erbB-2. This recombinant antibody can be used for  
XX the qualitative and quantitative determination of c-erbB-2. This can  
XX be used for monitoring or in-vivo localisation of tumours  
XX overexpressing c-erbB-2.

XX SQ Sequence 241 AA;  
Query Match 64.4%; Score 29; DB 13; Length 241;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
Db 134 gsdqltqs 142

RESULT 33  
AAR26981  
ID AAR26981 standard; Protein; 241 AA.  
XX AAR26981;  
XX  
DT 11-FEB-1993 (first entry)  
XX  
DE FWP51 fusion protein.  
XX  
KW Monoclonal antibody; light chain; heavy chain; tumour; c-erbB-2;  
KW variable region.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Domain 2..121 /label= FWP51\_heavy\_chain\_variable\_domain  
FT Region 32..36 /label= CDR1H  
FT Region 51..67 /label= CDR2H  
FT Region 100..110 /label= CDR3H  
FT Peptide 122..136 /label= Synthetic\_spacer  
FT Domain 137..241 /label= FWP52\_light\_chain\_variable\_domain  
FT Region 160..170 /label= CDR1L  
FT Region 186..192 /label= CDR2L  
FT Region 225..232 /label= CDR3L  
XX EP502812-A.  
PN  
XX  
XX 09-SEP-1992.  
XX  
XX 27-JAN-1992; 92EP-0810056.  
XX  
XX 05-FEB-1991; 91EP-0810079.  
XX  
XX (CIBA ) CIBA GEIGY AG.  
XX  
XX Groner B, Hardman N, Harwerth I, Hynes NE, Wells WS;  
XX Zwickl M;  
XX WPI; 1992-302096/37.  
XX N-PSDB; AAQ28260.  
XX  
XX Recombinant antibodies directed to growth factor receptor C-erbB-2 -  
XX for diagnosing and treating tumours expressing C-erbB-2 e.g. breast  
XX or ovarian tumours  
XX  
XX Disclosure; Page 43-45; 67pp; English.  
XX  
XX The sequence given is encoded by the single-chain Fv fusion gene  
XX comprising the monoclonal antibody FWP51 heavy and kappa light chain  
XX variable domains (see AAQ28258-9). This recombinant antibody can be used



CC for the qualitative and quantitative determination of c-erbB-2. This  
CC can be used for monitoring or in-vivo localisation of tumours  
CC overexpressing c-erbB-2.  
XX  
SQ Sequence 241 AA;  
  
Query Match 64.4%; Score 29; DB 13; Length 241;  
Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 GSIVQLSQS 10  
DB 135 gsdiglqtqs 143  
  
RESULT 34  
AAB31426  
ID AAB31426 standard; protein; 242 AA.  
XX  
AC AAB31426;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Protein used for immunising against shed antigen-specific B cells.  
XX  
KW Vaccine; shed antigen-specific B cell; idiotypic antibody;  
KW immune complex-mediated disease; autoimmune disease;  
KW humoral immune response; cancer.  
XX  
OS Synthetic.  
XX  
PN WO200076319-A1.  
XX  
PD 21-DEC-2000.  
XX  
PF 16-JUN-2000; 2000WO-US16677.  
XX  
PR 16-JUN-1999; 99US-0139521.  
PR 15-JUN-2000; 2000US-0594985.  
XX  
PA (BIOC-) BIOCRYSTAL LTD.  
XX  
XX Nelson MB, Barbera-Guillem E;  
PI  
DR WPI; 2001-080635/09.  
XX  
PT Inducing an immune response against shed antigen-specific B cell  
PT idiotypes, by administering a vaccine formulation comprising  
PT polynucleotides encoding an idiotypic determinant or peptides  
PT comprising an idiotypic determinant  
XX  
PS Example 2; Page 71; 81pp; English.  
XX  
CC The present sequence represents a synthetic protein which may be used  
CC in vaccines for inducing immune responses of the invention. The protein  
CC comprises a VH region linked to a VL region. The specification describes  
CC a method for inducing an immune response reactive with idiotypes present  
CC on shed antigen-specific B cells (SAB) of an individual. The method  
CC involves administering a vaccine formulation comprising polynucleotide  
CC encoding an idiotype of an antibody that binds to an epitope of shed  
CC antigen. The method is useful for inducing an immune response reactive  
CC with idiotypes present on SAB of an individual. The method is useful for  
CC depleting shed antigen-specific B cells, and for treating immune  
CC complex-mediated disease progression in organ specific autoimmune disease  
CC exacerbated by humoral immune response against groups expressed on shed  
CC antigen, or by plasma cell production of antibodies against groups of  
CC shed antigen. It is useful in cancer therapy and for treating autoimmune  
CC disease.  
XX  
SQ Sequence 242 AA;

Query Match 64.4%; Score 29; DB 22; Length 242;  
Best Local Similarity 77.8%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 GSIVQLSQS 10  
DB 123 gsgvqlqqs 131  
  
RESULT 35  
AAR77606  
ID AAR77606 standard; Protein; 248 AA.  
XX  
AC AAR77606;  
XX  
DT 15-MAR-1996 (first entry)  
XX  
DE Murine 5G1.1M1 scFv.  
XX  
KW Complement C5; haemolysis; kidney; glomerulonephritis;  
KW monoclonal antibody; antiinflammatory; antibody engineering;  
KW humanised antibody; complementarity determining region; CDR  
KW single chain antibody; scFv.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT Region 28..34  
FT /label= CDR-L1  
FT Region 52..54  
FT /label= CDR-L2  
FT Region 93..98  
FT /label= CDR-L3  
FT Region 156..159  
FT /label= CDR-H1  
FT Region 179..183  
FT /label= CDR-H2  
FT Region 226..236  
FT /label= CDR-H3  
XX  
PN WO9529697-A1.  
XX  
PD 09-NOV-1995.  
XX  
PF 01-MAY-1995; 95WO-US05688.  
XX  
PR 02-MAY-1994; 94US-0236208.  
XX  
PA (ALEX-) ALEXION PHARM INC.  
XX  
PI Evans MJ, Matis L, Mueller EE, Nye SH, Rollins S;  
PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;  
PI Wang Y, Wilkins JA;  
XX  
DR WPI; 1995-392923/50.  
DR N-PSDB; AAT08479.  
XX  
PT Treating glomerulonephritis with antibody against complement C5  
PT component - to inhibit complement induced cell lysis  
XX  
PS Example 11; Page 107-110; 181pp; English.  
XX  
CC Murine scFv 5G1.1M1 (AAR77606) is the product of a DNA construct  
CC (AAT08479) obt'd. from DNA of hybridoma ATCC HB 11625, which produces  
CC anti-C5 monoclonal antibody (Mab) 5G1.1. Recombinant scFv 5G1.1M1 is  
CC obt'd. by expression of this DNA in Escherichia coli using vector pET  
CC Trc S05/NI. The light and/or heavy chain CDRs of scFv 5G1.1M1 can  
CC be combined with CDRs from other 5G1.1-derived antibodies, Fds and  
CC light chains (AAR77607-16) in the prodn. of recombinant, including  
CC humanised, antibodies that retain the ability of Mab 5G1.1 to block  
CC human complement C5a generation and thus to reduce glomerular  
CC inflammation and kidney dysfunction associated with  
CC glomerulonephritis.

XX Sequence 248 AA; Query Match 64.4%; Score 29; DB 16; Length 248; Best Local Similarity 77.8%; Pred. No. 2e+02; Indels 2; Gaps 0; Matches 7; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

QY 2 GSIVQLSQS 10 Db 125 gsqvqlqqs 133

RESULT 36 AAW01829 AAW01829 standard; Protein; 248 AA. AAW01829; 20-JUL-1997 (first entry) scfv(1.1ASML) gene product. T-cell; receptor; CD44; epitope; fusion protein; cancer; immunoglobulin; ss. Synthetic. Location/Qualifiers 1..119 /label= VH 135..248 /label= VK 31..36 /label= CDR1 51..66 /label= CDR2 99..108 /label= CDR3 120..134 /label= linker 158..173 /label= CDR1 189..195 /label= CDR2 228..236 /label= CDR3 DE19540515-C1. 06-FEB-1997. 31-OCT-1995; 95DE-1040515. 31-OCT-1995; 95DE-1040515. (BOEH ) BOEHRINGER INGELHEIM INT GMBH. (GESL ) FORSCHUNGSZENTRUM KARLSRUHE GMBH. Hekele A, Herrlich P, Ponta H; WPI; 1997-101136/10. N-PSDB; AAT62573. T cells expressing fusion protein specific for variant CD44 gene - useful for cancer therapy, esp. pancreatic carcinoma Claim 7; Fig 1; 16pp; English. The fusion protein (AAT62573) comprises a first portion having specific affinity for an amino acid sequence encoded by a variant exon of the CD44 gene (pref. AAW01827 [rat] or AAW01828 [human]), and a second portion comprising at least part of the amino acid sequence of

CC a subunit of the T-cell receptor complex or of an immunoglobulin receptor. Products contg. the fusion protein (or DNA encoding it) are useful for treating cancer, esp. mammary, colonic, gastric or pancreatic carcinoma, and metastatic disease.

XX Sequence 248 AA; Query Match 64.4%; Score 29; DB 18; Length 248; Best Local Similarity 66.7%; Pred. No. 2e+02; Indels 1; Gaps 0; Matches 6; Conservative 2; Mismatches 1;

QY 2 GSIVQLSQS 10 Db 133 gsdqltqs 141

RESULT 37 AAW95668 AAW95668 standard; protein; 248 AA. AAW95668; 08-JUN-1999 (first entry) Mus musculus anti-IgE e27 sfv fragment. sfv fragment; IgE; antibody; anti-IgE; reduction; prevention; histamine; production; hypersensitivity; allergen; anaphylaxis; atopic allergy; asthma; allergic rhinitis; conjunctivitis; hay fever; eczema; anaphylactic shock; urticaria. Mus musculus. WO9901556-A2. 14-JAN-1999. 30-JUN-1998; 98WO-US13410. 02-JUL-1997; 97US-0887352. (GETH ) GENENTECH INC. Jardieu PM, Lowe J, Lowman HB, Presta LG; WPI; 1999-106057/09. Improving affinity of polypeptides, particularly anti-IgE antibodies - by identifying aspartyl residues which undergo isomerisation and substituting alternative residues and screening for affinity against the target Disclosure; Pages 103-104; 129pp; English. The sequence is that of the e27 sfv fragment. It was used as part of a method to improve the affinity of anti-IgE antibodies such as e26 and e27. The e26 and e27 antibodies can be used for reducing or preventing IgE mediated production of histamine in a mammal. They can be used for treating a disorder mediated by IgE such as hypersensitivity, atopic allergy, asthma, allergic rhinitis, conjunctivitis, hay fever, eczema, anaphylactic shock and urticaria. The antibodies can also be used for affinity purification, detection and diagnosis.

XX Sequence 248 AA; Query Match 64.4%; Score 29; DB 20; Length 248; Best Local Similarity 66.7%; Pred. No. 2e+02; Indels 1; Gaps 0; Matches 6; Conservative 2; Mismatches 1;

QY 2 GSIVQLSQS 10

Db 135 gsdiglts 143  
||:|:|  
RESULT 38  
AAW95667  
ID AAW95667 standard; protein; 248 AA.  
XX AC AAW95667;  
XX DT 08-JUN-1999 (first entry)  
XX DE Mus musculus anti-IgE e26 sFv fragment.  
XX KW sFv fragment; IgE; antibody; anti-IgE; reduction; prevention;  
KW histamine; production; hypersensitivity; allergen; anaphylaxis;  
KW atopic allergy; asthma; allergic rhinitis; conjunctivitis; hay fever;  
KW eczema; anaphylactic shock; urticaria.  
XX OS Mus musculus.  
XX PN W09901556-A2.  
XX PD 14-JAN-1999.  
XX PF 30-JUN-1998; 98WO-US13410.  
XX PR 02-JUL-1997; 97US-0887352.  
XX PA (GETH ) GENENTECH INC.  
XX PI Jardieu PM, Lowe J, Lowman HB, Presta LG;  
XX DR WPI; 1999-106057/09.  
XX PT Improving affinity of polypeptides, particularly anti-IgE antibodies  
PT - by identifying aspartyl residues which undergo isomerisation and  
PT substituting alternative residues and screening for affinity against  
PT the target  
XX PS Disclosure; Pages 102-103; 129pp; English.  
XX CC The sequence is that of the e26 sFv fragment. It  
CC was used as part of a method to improve the affinity of anti-IgE  
CC antibodies such as e26 and e27. The e26 and e27 antibodies can  
CC be used for reducing or preventing IgE mediated production of  
CC histamine in a mammal. They can be used for treating a disorder  
CC mediated by IgE such as hypersensitivity, atopic allergy, asthma,  
CC allergic rhinitis, conjunctivitis, hay fever, eczema, anaphylactic  
CC shock and urticaria. The antibodies can also be used for affinity  
CC purification, detection and diagnosis.  
XX SQ Sequence 248 AA;  
Query Match 64.4%; Score 29; DB 20; Length 248;  
Best Local Similarity 66.7%; Pred. NO. 2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GSIVQLSQS 10  
||:|:|  
Db 135 gsdiglts 143  
RESULT 39  
AAB76956  
ID AAB76956 standard; protein; 248 AA.  
XX AC AAB76956;  
XX DT 17-APR-2001 (first entry)  
XX KW sFv fragment of e26 SEQ ID 22.  
DE

XX KW Antibody; antiasthmatic; antiallergic; ophthalmological; dermatological;  
KW antiinflammatory; Ig E; immunoglobulin E; asthma; allergic rhinitis;  
KW conjunctivitis; eczema; urticaria; food allergy.  
XX OS Synthetic.  
XX PN US6172213-B1.  
XX PD 09-JAN-2001.  
XX PF 30-JUN-1998; 98US-0109207.  
XX PR 02-JUL-1997; 97US-0051554.  
XX PA (GETH ) GENENTECH INC.  
XX PI Lowman HB, Presta LG, Jardieu PM, Lowe J;  
XX DR WPI; 2001-122353/13.  
XX PT New nucleic acid encoding anti-immunoglobulin E antibody with improved  
PT properties, produced by substituting aspartyl residues in unimproved  
PT immunoglobulin E prone to isomerization by other residues by affinity  
PT maturation with phage display  
XX PS Disclosure; Fig 14; 87pp; English.  
XX CC This invention relates to a nucleotide sequence encoding an antibody  
CC with improved anti-IgE antibody activity. The antibody has improved  
CC action due to a process comprising, a) identifying aspartyl residues  
CC prone to isomerisation in unimproved anti-IgE (immunoglobulin E)  
CC antibody; b) substituting alternative residues to create candidate  
CC molecules, and c) selecting those candidate molecules which display  
CC affinity against the target molecule. Use of the antibody results in  
CC antiasthmatic; antiallergic; ophthalmological; dermatological and  
CC antiinflammatory activity. The antibodies are useful for treating  
CC IgE-mediated disorders such as asthma, allergic rhinitis, conjunctivitis,  
CC eczema, urticaria and food allergies. The mutant antibodies produced by  
CC the above mentioned nucleic acids may also be used as affinity  
CC purification agents and in diagnostic assays for detecting the expression  
CC of an antigen of interest in specific cell, tissues or serum. Amino acid  
CC sequences AAB76936-AAB76960 represent fragments of anti-IgE antibodies of  
CC the invention. Polynucleotide sequence AAF69253 represents an expression  
CC plasmid used in the course of the invention, and oligonucleotides  
CC AAF69254 - AAF69271 are used in the generation of affinity improved  
CC anti-IgE antibodies.  
XX SQ Sequence 248 AA;  
Query Match 64.4%; Score 29; DB 22; Length 248;  
Best Local Similarity 66.7%; Pred. NO. 2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GSIVQLSQS 10  
||:|:|  
Db 135 gsdiglts 143  
RESULT 40  
AAB76957  
ID AAB76957 standard; protein; 248 AA.  
XX AC AAB76957;  
XX DT 17-APR-2001 (first entry)  
XX DE sFv fragment of e27 SEQ ID 23.  
XX KW Antibody; antiasthmatic; antiallergic; ophthalmological; dermatological;  
KW antiinflammatory; Ig E; immunoglobulin E; asthma; allergic rhinitis;  
KW conjunctivitis; eczema; urticaria; food allergy.

XX OS Synthetic.  
XX PN US6172213-B1.  
XX PD 09-JAN-2001.  
XX PF 30-JUN-1998; 98US-0109207.  
XX PR 02-JUL-1997; 97US-0051554.  
XX PA (GETH ) GENENTECH INC.  
XX PI Lowman HB, Presta LG, Jardieu PM, Lowe J;  
XX DR WPI; 2001-122353/13.  
XX  
XX New nucleic acid encoding anti-immunoglobulin E antibody with improved properties, produced by substituting aspartyl residues in unimproved immunoglobulin E prone to isomerization by other residues by affinity maturation with phage display .  
XX Claim 3; Fig 14; 87pp; English.  
XX  
XX This invention relates to a nucleotide sequence encoding an antibody with improved anti-IgE antibody activity. The antibody has improved action due to a process comprising, a) identifying aspartyl residues prone to isomerization in unimproved anti-IgE (immunoglobulin E) antibody; b) substituting alternative residues to create candidate molecules, and c) selecting those candidate molecules which display affinity against the target molecule. Use of the antibody results in anti-inflammatory activity. The antibodies are useful for treating allergic disorders such as asthma, allergic rhinitis, conjunctivitis, eczema, urticaria and food allergies. The mutant antibodies produced by the above mentioned nucleic acids may also be used as affinity purification agents and in diagnostic assays for detecting the expression of an antigen of interest in specific cell, tissues or serum. Amino acid sequences AAB76936-AAB76960 represent fragments of anti-IgE antibodies of the invention. Polynucleotide sequence AAF69253 represents an expression plasmid used in the course of the invention, and oligonucleotides AAF69254 - AAF69271 are used in the generation of affinity improved anti-IgE antibodies.  
XX Sequence 248 AA;  
XX  
XX Query Match 64.4%; Score 29; DB 22; Length 248;  
XX Best Local Similarity 66.7%; Pred. No. 2e+02;  
XX Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX QY 2 GSVVQLSQS 10  
XX II :II:II  
XX Db 135 gsdqiltqs 143  
XX  
XX RESULT 41  
XX AAP80152 standard; protein; 251 AA.  
XX AC AAP80152;  
XX DT 20-OCT-1990 (first entry)  
XX DE Multifunctional protein.  
XX KW Multifunctional protein; biosynthetic antibody binding site;  
XX KW murine monoclonal 26-10;  
XX PN WO8809344-A.  
XX PD 01-DEC-1988.  
XX

PF 19-MAY-1988; 88WO-US01737.  
PR 21-MAY-1987; 87US-0052800.  
XX (CREA-) CREATIVE BIOMOLECULES INC.  
XX Huston JS, Oppermann H;  
XX WPI; 1988-353928/49.  
XX N-PSDB; AAN80178.  
XX  
XX Recombinant multifunctional protein - having antibody binding site and a sequence for biological activity, ion sequestering or binding to a solid support.  
XX Disclosure; : 15pp; English.  
XX  
XX The multifunctional protein comprises a single chain biosynthetic Ab binding site (comprising sequences mimicking the light chain and heavy chain variable regions of murine monoclonal 26-10) with the specificity of murine monoclonal 26-10, linked through a spacer to the FB fragment of protein A, here fused as a leader, and constituting a binding site for Fc The spacer contains the 11 C-terminal amino acids of the FB followed by Asp-Pro ( a dilute acid cleavage site). The construct contains binding sites for both Fc and digoxin, and its structure can be summarised as: FB-Asp-Pro-VL-(Gly4-Ser)3-VH. The protein can be used for specific binding assays, affinity purification, biocatalysts, targeting, imaging an immunological treatment of oncogenic and infectious diseases. It offers fewer cleavage sites to circulating proteolytic enzymes, and improved stability. It reaches target tissues rapidly and is quickly cleared from the body. It also has reduced immunogenicity and its design facilitates binding to other moieties in drug targeting and imaging applications.  
XX Sequence 251 AA;  
XX  
XX Query Match 64.4%; Score 29; DB 9; Length 251;  
XX Best Local Similarity 77.8%; Pred. No. 2e+02;  
XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX QY 2 GSVVQLSQS 10  
XX II :II:II  
XX Db 131 gsevqltqs 139  
XX  
XX RESULT 42  
XX AAB28159  
XX ID AAB28159 standard; protein; 253 AA.  
XX AC AAB28159;  
XX DT 08-FEB-2001 (first entry)  
XX DE Murine anti-EGP-2 single chain Fv fragment.  
XX KW Murine; immunoglobulin; antigen-binding; framework region; carcinoma;  
XX KW c-erbB2; carcinoma-associated antigen.  
XX OS Mus sp.  
XX PN WO200061635-A2.  
XX PD 19-OCT-2000.  
XX PF 10-APR-2000; 2000WO-EP03176.  
XX PR 09-APR-1999; 99EP-0107030.  
XX PA (UYZU-) UNIV ZUERICH.  
XX PA (PLUE/) PLUECKTHUN A.  
XX PI Plueckthun A, Honegger A, Willuda J;



XX WPI; 2000-679468/66.

XX Stabilizing chimeric immunoglobulin (Ig) involves setting up a

PT stabilized antigen binding Ig or its fragment by replacing one or more

PT residues present in acceptor Ig by those residues present in donor Ig

PT

XX Claim 6; Page 51; 51pp; English.

XX The present invention relates to a method for stabilising a chimeric

CC immunoglobulin (Ig). The method comprises identifying antigen-binding

CC groups derived from donor Ig and framework regions derived from an

CC acceptor Ig. The present sequence is one such donor Ig fragment.

CC One or more of the residues present at the positions in the present

CC sequence are replaced by those present at the corresponding positions in

CC the donor Ig, after comparing the structural features of the VH domains

CC of the acceptor Ig and the donor Ig. The acceptor Ig fragment used in

CC the present invention is human anti-c-ErbB2 single chain Fv fragment 4D5

CC (AAB28158). The method of the present invention is useful for producing

CC a pharmaceutical composition which can be used for treating human

CC carcinomas, since c-erbB2 is a carcinoma-associated antigen.

XX

SQ Sequence 253 AA;

Query Match 64.4%; Score 29; DB 21; Length 253;

Best Local Similarity 77.8%; Pred. No. 2e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

Db 136 gsqvqlqgs 144

RESULT 43

AAR20185

ID AAR20185 standard; peptide; 271 AA.

XX

AC AAR20185;

XX

DT 15-APR-1992 (first entry)

XX

DE Sequence of the shortened hinge version of the B72.3 single

DE chain Fv hinge.

XX

KW Fv fragment; in vivo diagnosis; therapy; antibody.

XX

OS Homo sapiens.

XX

PN WO9119739-A.

XX

PD 26-DEC-1991.

XX

PF 11-JUN-1991; 91WO-GB00935.

XX

PR 11-JUN-1990; 90GB-0012995.

XX

PA (CELL-) CELLTECH LTD.

XX

PI King DJ, Mountain A, Owens RJ, Yarranton GT;

XX

DR WPI; 1992-024365/03.

DR N-PSDB; AAQ20381.

XX

PT New multivalent antigen-binding proteins - comprise Fv fragment

PT linked to at least 1 other Fv fragment spacer and useful for

PT in-vivo diagnosis or therapy

XX

PS Example; Fig 5; 54pp; English.

XX

CC The bivalent antigen-binding protein B72.3 Fv (AAR20184) is an example

CC of the antigen-binding proteins of the invention. They comprise a

CC first Fv fragment bound to at least one other Fv fragment by a

CC linker which keeps the Fv fragment apart. The connecting structure

CC of AAR20184 comprises a joining sequence derived from a human IGG1

CC domain linked to a complete human IGG4 hinge region. AAR20185 is the

CC sequence of the shortened hinge version of the B72.3 single chain Fv

CC hinge.

XX

SQ Sequence 271 AA;

Query Match 64.4%; Score 29; DB 13; Length 271;

Best Local Similarity 77.8%; Pred. No. 2.2e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

Db 144 gsqvqlqgs 152

RESULT 44

AAR34672

ID AAR34672 standard; Protein; 272 AA.

XX

AC AAR34672;

XX

DT 25-AUG-1993 (first entry)

XX

DE 26-10 sFv.

XX

KW Heavy; light; variable; VH; VL; region; antidigitoxin; monoclonal;

KW antibody; MAb; 26-10; Fv; fragment; antigen; binding site; linker;

KW expression-secretion system; T7 promoter; signal peptide; molecule;

KW polymerase chain reaction; PCR; single chain.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 1..25

FT /note= "Signal peptide"

FT Region 26..138

FT /note= "26-10 VL sequence"

FT Peptide 139..153

FT /label= Linker peptide

FT Region 154..272

FT /note= "26-10 VH sequence"

XX

PN WO9308300-A.

XX

PD 29-APR-1993.

XX

PF 16-OCT-1992; 92WO-US08881.

XX

PR 18-OCT-1991; 91US-0777709.

XX

PA (UYCA-) UNIV CALGARY.

XX

PI Anthony JG, Ng SC, Wong S;

XX

DR WPI; 1993-152491/18.

DR N-PSDB; AAQ41069.

XX

PT Expression-secretion vectors - for prodn. of biologically active

PT antibody Fv fragments or single chain Fv molecules

XX

PS Disclosure; Fig 6; 51pp; English.

XX

CC This sequence represents a single chain Fv molecule which contains

CC the heavy and light variable portions (VH and VL) of antidigitoxin

CC monoclonal antibody (MAb) 26-10 (see also AAR34670-71), a signal

CC sequence (see AAR38527-28) and a linker between the two variable

CC regions. The DNA sequence encoding this protein was used in the

CC construction of an expression-secretion system for the production

CC of biologically active Fv fragments. The system also contains a

CC DNA sequence encoding the T7 promoter. The secretion-expression  
CC vector was produced by polymerase chain reaction Fv fragments  
CC comprising only the VH or VL regions, and single chain molecules may  
CC be used to specifically bind one or more of the same antigens as the  
CC full length antibody from which they are derived.  
XX  
SQ Sequence 272 AA;

Query Match 64.4%; Score 29; DB 14; Length 272;  
Best Local Similarity 77.8%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
II III II  
Db 152 gsevqlqqs 160

RESULT 45  
AAW00557  
ID AAW00557 standard; Protein; 272 AA.  
XX  
AC AAW00557;

XX 31-OCT-1996 (first entry)

DT Nematode salivary protein antibody construct.

DE Cyst nematode; salivary protein; svp49; transgenic plant;  
KW biological control; crop protection; antibody engineering;  
KW Heterodera; Globodera; Punctodera; monoclonal antibody.

XX Mus musculus.

XX Key Location/Qualifiers  
FH Domain 1..112  
FT /label= Light\_chain\_variable\_domain  
FT 25..39  
FT /label= CDR  
FT 55..61  
FT /label= CDR  
FT 94..102  
FT /label= CDR  
FT 113..127  
FT /label= Gly-Ser-linker  
FT 128..272  
FT /label= Heavy\_chain\_variable\_domain  
FT 158..162  
FT /label= CDR  
FT 177..193  
FT /label= CDR  
FT 226..232  
FT /label= CDR

XX WO9622372-A2.

PN 25-JUL-1996.

PD 17-JAN-1996; 96WO-NL00033.

XX 17-JAN-1995; 95EP-0200110.

XX (RIJK-) RIJKS LANDBOUWHOGESCHOOL WAGENINGEN.

XX Bakker J, Schots A, Stiekema WJ;

XX WPI; 1996-354529/35.

DR N-PSDB; AAT35691.

XX Antibody against stage two juvenile nematode salivary proteins  
PT used in the production of transgenic plants, resistant to nematode  
PT infection  
XX

PS Claim 7; Page 17-19; 25pp; English.

XX An engineered antibody construct (AAW00557) comprises the 2 variable  
CC regions of an antibody that specifically reacts with a salivary  
CC protein of cyst nematodes, fused by a 15-mer linker peptide. The  
CC antibody is pref. mouse monoclonal antibody M848, which recognises  
CC subventral oesophageal proteins svp30, svp31a, svp31b, svp32, svp39  
CC and svp49 (see also AAW00556) of cyst nematode second stage juveniles  
CC (J2), and which was raised using J2 protein fractions of Globodera  
CC rostochiensis. Transgenic plants expressing such antibody  
CC constructs are protected against attack from Heterodera, Punctodera  
CC and Globodera cyst nematodes.

XX Sequence 272 AA;

Query Match 64.4%; Score 29; DB 17; Length 272;  
Best Local Similarity 77.8%; Pred. No. 2.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
II III II  
Db 126 gsevqlqqs 134

Search completed: June 28, 2001, 11:54:19  
Job time: 80 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:04 ; Search time 42.81 Seconds  
(without alignments)  
17.794 Million cell updates/sec

Title: US-09-439-313-573  
Perfect score: 45  
Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	73.3	3413	2 T17467	rifamycin polyketi
2	31	68.9	304	2 T03597	phosphoprotein pho
3	31	68.9	394	2 T24860	hypothetical prote
4	31	68.9	436	2 D83498	C4-dicarboxylate t
5	31	68.9	468	2 A49289	peroxisome prolif
6	31	68.9	468	2 JC2085	peroxisome prolif
7	31	68.9	468	2 A45288	peroxisome prolif
8	31	68.9	468	2 I56603	hypothetical prote
9	31	68.9	617	2 T46319	adenylate cyclase
10	31	68.9	838	2 A38172	sucrose-phosphate
11	30	66.7	117	2 S53083	transcription elon
12	30	66.7	161	2 C75331	ribosomal protein
13	30	66.7	193	2 S78130	probable short-cha
14	30	66.7	241	2 C83133	probable flagellar
15	30	66.7	280	2 G81421	hypothetical prote
16	30	66.7	305	2 S32237	GRR1 related prote
17	30	66.7	364	2 T48752	probable transport
18	30	66.7	382	2 E85615	membrane protein y
19	30	66.7	382	2 A64829	hypothetical prote
20	30	66.7	425	2 T08946	carboxypeptidase r
21	30	66.7	503	2 E82699	hypothetical prote
22	30	66.7	539	2 B72707	C4-dicarboxylate s
23	30	66.7	557	2 S30288	hypothetical prote
24	30	66.7	662	2 T17339	sensory transducti
25	30	66.7	750	2 S75636	DNA-directed RNA p
26	30	66.7	880	1 B33926	hypothetical prote
27	30	66.7	964	2 A84693	hypothetical prote
28	30	66.7	968	2 F84693	hypothetical prote
29	30	66.7	972	2 E84693	hypothetical prote

30	66.7	1059	2 T12195	sucrose-phosphate
31	66.7	1277	2 S53043	probable membrane
32	66.7	1341	2 T17285	hypothetical prote
33	66.7	1520	2 T23620	hypothetical prote
34	66.7	2559	2 T09144	probable guanine n
35	66.7	5069	2 T17464	rifamycin polyketi
36	66.7	188	2 S12393	hypothetical prote
37	66.7	225	2 G86748	hypothetical prote
38	66.7	254	2 F84145	transposase (23) B
39	66.7	261	2 H84149	transposase (27) B
40	66.7	274	2 T29574	hypothetical prote
41	66.7	285	2 H71938	flagellar motor sw
42	66.7	287	2 F64648	fliY protein - Hel
43	66.7	305	2 A40573	clathrin heavy cha
44	66.7	340	2 H70790	hypothetical prote
45	66.7	378	2 A85877	probable prophage
46	66.7	380	2 T19096	hypothetical prote
47	66.7	384	2 S64074	hypothetical prote
48	66.7	459	1 S04904	glutamate dehydrog
49	66.7	463	2 A65159	seryl-tRNA(Sec) se
50	66.7	463	2 C86034	hypothetical prote
51	66.7	503	1 QQXPGP	surface glycoprote
52	66.7	504	2 T50270	hypothetical prote
53	66.7	511	2 S47860	amidophosphoribosy
54	66.7	540	2 S41864	kinesin light chai
55	66.7	563	2 S47997	kinesin light chai
56	66.7	573	2 JQ0135	hypothetical 62.8K
57	66.7	582	2 T39931	probable transcrip
58	66.7	610	2 H83243	probable ATP-bindi
59	66.7	622	1 A49344	cell wall assembly
60	66.7	679	2 I48084	gibbon ape leukemi
61	66.7	958	2 T10679	hypothetical prote
62	66.7	1171	2 T17454	diaphanous-related
63	66.7	1262	2 T30524	protein phosphatas
64	66.7	1569	2 G02757	clathrin heavy cha
65	66.7	1640	2 T09522	clathrin heavy cha

ALIGNMENTS

RESULT 1  
T17467  
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei  
C;Species: Amycolatopsis mediterranei  
C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 17-Nov-2000  
C;Accession: T17467  
R;Schupp, T.  
submitted to the EMBL Data Library, December 1997  
A;Reference number: Z18802  
A;Accession: T17467  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-3413 <SCH>  
A;Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1  
A;Experimental source: strain LBG A3136  
C;Superfamily: acyl carrier protein homology  
C;Keywords: carrier protein  
F;1608-1679/Domain: acyl carrier protein homology <ACP1>  
F;3334-3405/Domain: acyl carrier protein homology <ACP2>  
  
Query Match 73.3%; Score 33; DB 2; Length 3413;  
Best Local Similarity 60.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:|::||| |  
Db 2588 LGAVVQLPQS 2597

RESULT 2  
T03597



phosphoprotein phosphatase (EC 3.1.3.16) 1, npp3 - common tobacco  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
 C:Accession: T03597  
 R:Suhr, M.; Cho, H.; Kim, Y.; Liu, J.; Lee, H.  
 Plant Mol. Biol. 36, 315-322, 1998  
 A:Title: Multiple genes encoding serine/threonine protein phosphatases and their differ  
 A:Reference number: Z14967; MUID:98145437  
 A:Accession: T03597  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-304 <SUH>  
 A:Cross-references: EMBL:293770; NID:q1929407; PIDN:CAB07805.1; PID:q1929408  
 A:Experimental source: cultivar Xanthi.  
 C:Genetics:  
 A:Note: npp3  
 C:Superfamily: phosphoprotein phosphatase; phosphoesterase core homology; phosphoprotein  
 C:Keywords: iron; phosphoric monoester hydrolase; serine/threonine-specific phosphatase;  
 F:27-286/Domain: phosphoprotein phosphatase homology <PPP>  
 F:55-123/Domain: phosphoesterase core homology <PEC>

Query Match 68.9%; Score 31; DB 2; Length 304;  
 Best Local Similarity 66.7%; Pred. NO. 39;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
 I:|||||  
 Db 22 GKLVLSES 30

RESULT 3  
 T24860  
 hypothetical protein T12D8.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T24860  
 R:McMurray, A.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19944  
 A:Accession: T24860  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-394 <WIL>  
 A:Cross-references: EMBL:Z81120; PIDN:CAB03344.1; GSPDB:GN00021; CESP:T12D8.4  
 A:Experimental source: clone T12D8  
 C:Genetics:  
 A:Gene: CESP:T12D8.4  
 A:Map position: 3  
 A:Introns: 27/3; 58/1; 102/3; 159/3; 205/2; 289/3; 346/3

Query Match 68.9%; Score 31; DB 2; Length 394;  
 Best Local Similarity 77.8%; Pred. NO. 51;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
 :||| ||||  
 Db 302 IGSIVQLSQ 310

RESULT 4  
 D83498  
 C4-dicarboxylate transport protein PA1183 [imported] - Pseudomonas aeruginosa (strain PA  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: D83498  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337

A:Accession: D83498  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-436 <STO>  
 A:Cross-references: GB:AE004548; GB:AE004091; NID:g9947102; PIDN:AAG04572.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: dctA; PA1183  
 C:Superfamily: C4-dicarboxylate carrier protein

Query Match 68.9%; Score 31; DB 2; Length 436;  
 Best Local Similarity 66.7%; Pred. NO. 57;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
 :||| |||  
 Db 214 VGSIVQLSQ 222

RESULT 5  
 A49289  
 peroxisome proliferator-activated receptor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 20-Sep-1999  
 C:Accession: A49289  
 R:Sherr, T.; Yi, H.F.; McBride, O.W.; Gonzalez, F.J.  
 Biochemistry 32, 5598-5604, 1993  
 A:Title: cDNA cloning, chromosomal mapping, and functional characterization of the hu  
 A:Reference number: A49289; MUID:93277839  
 A:Accession: A49289  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-468 <SHE>  
 A:Cross-references: GB:L02932; NID:g307340; PIDN:AAA36468.1; PID:g307341  
 A:Note: authors translated the codon AAC for residue 33 as Asp  
 C:Genetics:  
 A:Map position: 22  
 C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
 C:Keywords: DNA binding; transcription regulation; zinc finger  
 F:100-386/Domain: erba transforming protein homology <ERBA>

Query Match 68.9%; Score 31; DB 2; Length 468;  
 Best Local Similarity 60.0%; Pred. NO. 61;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 ||| :|||  
 Db 31 MGNIQEISQS 40

RESULT 6  
 JC2085  
 peroxisome proliferator-activated receptor alpha chain - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 20-Sep-1999  
 C:Accession: JC2085; S11659; S68199; S67963  
 R:Gearring, K.L.; Crickmore, A.; Gustafsson, J.A.  
 Biochem. Biophys. Res. Commun. 199, 255-263, 1994  
 A:Title: Structure of the mouse peroxisome proliferator activated receptor alpha gene  
 A:Reference number: JC2085; MUID:94168583  
 A:Accession: JC2085  
 A:Molecule type: DNA  
 A:Residues: 1-468 <GEA>  
 A:Cross-references: EMBL:X75287; EMBL:X75288; EMBL:X75289; NID:g467773; PIDN:CRA53042  
 R:Issemann, I.; Green, S.  
 Nature 347, 645-650, 1990  
 A:Title: Activation of a member of the steroid hormone receptor superfamily by peroxi  
 A:Reference number: S11659; MUID:91015382  
 A:Accession: S11659  
 A:Molecule type: mRNA  
 A:Residues: 1-74, 'R', 76-468 <ISS>

A;Cross-references: GB:X57638; NID:g53764; PIDN:CAA40856.1; PID:g53765  
R;Jones, P.S.; Savory, R.; Barratt, P.; Bell, A.R.; Gray, T.J.B.; Jenkins, N.A.; Gilbert  
Eur. J. Biochem. 233, 219-226, 1995  
A;Title: Chromosomal localisation, inducibility, tissue-specific expression and strain d  
A;Reference number: S67963; MUID:96061953  
A;Accession: S68199  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 413-468 <JON>  
A;Cross-references: EMBL:X89577; NID:g1051294; PIDN:CAA61754.1; PID:g1051295  
A;Experimental source: strain Swiss Webster; tissue liver  
A;Accession: S67963  
A;Molecule type: protein  
A;Residues: 1-5 <JOW>  
C;Comment: This protein mediates transcription of responsive genes which include those e  
C;Genetics:  
A;Gene: PPARalpha  
A;Map position: 15  
A;Introns: 70/1; 123/2; 170/1; 237/3; 387/1  
C;Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C;Keywords: DNA binding; receptor; transcription regulation; zinc finger  
F;100-386/Domain: erba transforming protein homology <ERBA>  
F;102-165/Region: DNA binding  
F;139-156/Region: zinc finger CCCC motif  
F;273-468/Domain: ligand binding #status predicted <LIB>

Query Match 68.9%; Score 31; DB 2; Length 468;  
Best Local Similarity 60.0%; Pred. No. 61;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
||:| :|||  
Db 31 MGNIQEISQS 40

RESULT 7  
A45288  
peroxisome proliferator-activated receptor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 20-Sep-1999  
C;Accession: A45288  
R;Goettlicher, M.; Widmark, E.; Li, Q.; Gustafsson, J.A.  
Proc. Natl. Acad. Sci. U.S.A. 89, 4653-4657, 1992  
A;Title: Fatty acids activate a chimera of the clofibric acid-activated receptor and the  
A;Reference number: A45288; MUID:92262498  
A;Accession: A45288  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-468 <GOE>  
A;Cross-references: GB:M88592; NID:g206317; PIDN:AAA41918.1; PID:g206318  
C;Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C;Keywords: DNA binding; transcription regulation; zinc finger  
F;100-386/Domain: erba transforming protein homology <ERBA>

Query Match 68.9%; Score 31; DB 2; Length 468;  
Best Local Similarity 60.0%; Pred. No. 61;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGSIVQLSQS 10  
||:| :|||  
Db 31 MGNIQEISQS 40

RESULT 8  
I56603  
peroxisome proliferator activated receptor alpha - human  
C;Species: Homo sapiens (man)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Sep-1999  
C;Accession: I56603  
R;Mukherjee, R.; Jow, L.; Noonan, D.; McDonnell, D.P.  
J. Steroid Biochem. Mol. Biol. 51, 157-166, 1994

A;Title: Human and rat peroxisome proliferator activated receptors (PPARS) demonstrat  
A;Reference number: I56603; MUID:95071923  
A;Accession: I56603  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-468 <RES>  
A;Cross-references: GB:S74349; NID:g765239; PIDN:AAB32649.1; PID:g765240  
C;Genetics:  
A;Gene: hPPARalpha  
C;Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C;Keywords: zinc finger  
F;100-386/Domain: erba transforming protein homology <ERBA>

Query Match 68.9%; Score 31; DB 2; Length 468;  
Best Local Similarity 60.0%; Pred. No. 61;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
||:| :|||  
Db 31 MGNIQEISQS 40

RESULT 9  
T46319  
hypothetical protein DKF2p434G0812.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C;Accession: T46319  
R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A;Reference number: Z23035  
A;Accession: T46319  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-617 <AAA>  
A;Cross-references: EMBL:AL137619  
A;Experimental source: adult testis; clone DKF2p434G0812  
C;Genetics:  
A;Note: DKF2p434G0812.1

Query Match 68.9%; Score 31; DB 2; Length 617;  
Best Local Similarity 60.0%; Pred. No. 82;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:|:| :|||  
Db 275 LGSLEQLSQA 284

RESULT 10  
A38172  
adenylate cyclase (EC 4.6.1.1) - Pasteurella multocida  
C;Species: Pasteurella multocida  
C;Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 24-Nov-1999  
C;Accession: A38172  
R;Mock, M.; Crasnier, M.; Duflot, E.; Dumay, V.; Danchin, A.  
J. Bacteriol. 173, 6265-6269, 1991  
A;Title: Structural and functional relationships between Pasteurella multocida and en  
A;Reference number: A38172; MUID:92011391  
A;Accession: A38172  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-838 <MOC>  
A;Cross-references: GB:M68901; NID:g150497; PIDN:AAA25532.1; PID:g551912  
C;Superfamily: adenylate cyclase  
C;Keywords: phosphorus-oxygen lyase

Query Match 68.9%; Score 31; DB 2; Length 838;  
Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 |||||:|:  
 Db 104 MGSIASISQT 113

RESULT 11  
 S53083  
 sucrose-phosphate synthase - fava bean (fragment)  
 C:Species: Vicia faba (fava bean)  
 C:Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 18-Jun-1999  
 C:Accession: S53083  
 R: Buchner, P.  
 submitted to the EMBL Data Library, March 1995  
 A:Description: Vicia faba sucrose-phosphate synthase cDNA fragment.  
 A:Reference number: S53083  
 A:Accession: S53083  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-117 <BUC>  
 A:Cross-references: EMBL:Z48640; NID:g732985; PIDN:CAA88587.1; PID:g732986  
 C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology  
 F:1-117/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SSPS>

Query Match 66.7%; Score 30; DB 2; Length 117;  
 Best Local Similarity 50.0%; Pred. No. 25;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 |||||:|:  
 Db 89 MGHIQMSKA 98

RESULT 12  
 C75331  
 transcription elongation factor - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: C75331  
 R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: C75331  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-161 <WHI>  
 A:Cross-references: GB:AE002035; GB:AE000513; NID:g6459753; PIDN:AAF11523.1; PID:g645975

A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR1970  
 A:Map position: 1  
 C:Superfamily: transcription elongation factor greb  
 C:Keywords: transcription factor

Query Match 66.7%; Score 30; DB 2; Length 161;  
 Best Local Similarity 75.0%; Pred. No. 34;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLS 8  
 :|:|||||  
 Db 88 LGAIVQLS 95

RESULT 13  
 S78130  
 ribosomal protein L10 - Reclinomonas americana (ATCC 50394) mitochondrion  
 C:Species: mitochondrion Reclinomonas americana  
 A:Variety: ATCC 50394

C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 05-Nov-1999  
 C:Accession: S78130  
 R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; S  
 Nature, 387, 493-497, 1997  
 A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.  
 A:Reference number: S78127; MUID:97311393

A:Accession: S78130  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-193 <LAN>  
 A:Cross-references: EMBL:AF007261; NID:g2258325; PIDN:AAD11863.1; PID:g2258329  
 A:Experimental source: AFCC 50394  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997  
 C:Genetics:  
 A:Gene: rpl10  
 A:Genome: mitochondrion  
 A:Start codon: GTG  
 C:Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match 66.7%; Score 30; DB 2; Length 193;  
 Best Local Similarity 75.0%; Pred. No. 41;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SIVQLSQS 10  
 |:|||||:  
 Db 161 SVVQLTQS 168

RESULT 14  
 C83133  
 probable short-chain dehydrogenase PA4098 [imported] - Pseudomonas aeruginosa (strain  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: C83133  
 R:Stover, C.K.; Pham, X.-Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: C83133  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-241 <STO>  
 A:Cross-references: GB:AE004826; GB:AE004091; NID:g9950296; PIDN:RAG07485.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4098  
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 66.7%; Score 30; DB 2; Length 241;  
 Best Local Similarity 66.7%; Pred. No. 52;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
 |:|||||:  
 Db 151 GAIVQLTRS 159

RESULT 15  
 G81421  
 probable flagellar motor switch protein Cj0059c [imported] - Campylobacter jejuni (st  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000  
 C:Accession: G81421  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil  
 C.W.; Quail, M.; Rajadream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
 A:Reference number: A81250; MUID:20150912  
 A:Accession: G81421

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-280 <PAR>  
A;Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72546.1; PID:g6967505  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: fliv; Cj0059c

Query Match 66.7%; Score 30; DB 2; Length 280;  
Best Local Similarity 55.6%; Pred. No. 60;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:||:|:|:  
Db 227 IGSVVELNQ 235

RESULT 16  
S32237  
hypothetical protein 31 (rplJ 5' region) - Streptomyces griseus  
C;Species: Streptomyces griseus  
C;Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 10-Mar-1994  
C;Accession: S32237  
R;Kuester, K.; Kuberski, S.; Piepersberg, W.; Distler, J.  
submitted to the EMBL Data Library, March 1993  
A;Description: Cloning and nucleotide sequence analysis of the nusG-rplK-rplA-rplJ-rplL  
A;Reference number: S32234  
A;Accession: S32237  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-305 <KUE>  
A;Cross-references: EMBL:X72787

Query Match 66.7%; Score 30; DB 2; Length 305;  
Best Local Similarity 50.0%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

QY 1 MGSIVQLSQS 10  
|||:::|:  
Db 75 MGSVMSMKQS 84

RESULT 17  
T48752  
GRR1 related protein [imported] - Neurospora crassa (fragment)  
N;Alternate names: protein 8D4.260  
C;Species: Neurospora crassa  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
C;Accession: T48752  
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24541  
A;Accession: T48752  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-364 <SCH>  
A;Cross-references: EMBL:AL353819; GSPDB:GN00112; NCSP:8D4.260  
A;Experimental source: cosmid contig 8D4; strain 74  
C;Genetics:  
A;Gene: NCSP:8D4.260  
A;Map position: 2  
A;Introns: 267/3

Query Match 66.7%; Score 30; DB 2; Length 364;  
Best Local Similarity 60.0%; Pred. No. 79;  
Matches 6; Conservative 3; Mismatches 1; Indels 1; Gaps 0;

QY 1 MGSIVQLSQS 10  
:|:|:|:|:  
Db 252 IASLVQLAQs 261

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-280 <PAR>  
A;Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72546.1; PID:g6967505  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: fliv; Cj0059c

Query Match 66.7%; Score 30; DB 2; Length 280;  
Best Local Similarity 55.6%; Pred. No. 60;  
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:||:|:|:  
Db 227 IGSVVELNQ 235

RESULT 16  
S32237  
hypothetical protein 31 (rplJ 5' region) - Streptomyces griseus  
C;Species: Streptomyces griseus  
C;Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 10-Mar-1994  
C;Accession: S32237  
R;Kuester, K.; Kuberski, S.; Piepersberg, W.; Distler, J.  
submitted to the EMBL Data Library, March 1993  
A;Description: Cloning and nucleotide sequence analysis of the nusG-rplK-rplA-rplJ-rplL  
A;Reference number: S32234  
A;Accession: S32237  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-305 <KUE>  
A;Cross-references: EMBL:X72787

Query Match 66.7%; Score 30; DB 2; Length 305;  
Best Local Similarity 50.0%; Pred. No. 66;  
Matches 5; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

QY 1 MGSIVQLSQS 10  
|||:::|:  
Db 75 MGSVMSMKQS 84

RESULT 17  
T48752  
GRR1 related protein [imported] - Neurospora crassa (fragment)  
N;Alternate names: protein 8D4.260  
C;Species: Neurospora crassa  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
C;Accession: T48752  
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24541  
A;Accession: T48752  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-364 <SCH>  
A;Cross-references: EMBL:AL353819; GSPDB:GN00112; NCSP:8D4.260  
A;Experimental source: cosmid contig 8D4; strain 74  
C;Genetics:  
A;Gene: NCSP:8D4.260  
A;Map position: 2  
A;Introns: 267/3

Query Match 66.7%; Score 30; DB 2; Length 364;  
Best Local Similarity 60.0%; Pred. No. 79;  
Matches 6; Conservative 3; Mismatches 1; Indels 1; Gaps 0;

QY 1 MGSIVQLSQS 10  
:|:|:|:|:  
Db 252 IASLVQLAQs 261

RESULT 18  
E85615  
probable transport ycaD [imported] - Escherichia coli (strain O157:H7)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: E85615  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
illier, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: E85615  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-382 <STO>  
A;Cross-references: GB:AE005174; NID:g12514063; PIDN:AAG55385.1; GSPDB:GN00145; UWGP:  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: ycaD

Query Match 66.7%; Score 30; DB 2; Length 382;  
Best Local Similarity 60.0%; Pred. No. 83;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:|:|:|:|:  
Db 277 LGSIAMLSQA 286

RESULT 19  
A64829  
membrane protein ycaD - Escherichia coli  
C;Species: Escherichia coli  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C;Accession: A64829; S03788  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617  
A;Accession: A64829  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-382 <BLAT>  
A;Cross-references: GB:AE000192; GB:U00096; NID:g1787125; PIDN:AAC73984.1; PID:g17871  
A;Experimental source: strain K-12, substrain MGL1655  
R;Bilous, P.T.; Cole, S.T.; Anderson, W.F.; Weiner, J.H.  
Mol. Microbiol. 2, 785-795, 1988  
A;Title: Nucleotide sequence of the dmsABC operon encoding the anaerobic dimethylsulp  
A;Reference number: S03784; MUID:89096500  
A;Accession: S03788  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-264,'H' <BIL>  
C;Genetics:  
A;Gene: ycaD  
A;Map position: 20 min  
C;Keywords: transmembrane protein  
F;8-24/Domain: transmembrane #status predicted <TM1>  
F;76-92/Domain: transmembrane #status predicted <TM2>  
F;103-119/Domain: transmembrane #status predicted <TM3>  
F;160-176/Domain: transmembrane #status predicted <TM4>  
F;209-225/Domain: transmembrane #status predicted <TM5>  
F;235-251/Domain: transmembrane #status predicted <TM6>  
F;265-281/Domain: transmembrane #status predicted <TM7>  
F;286-302/Domain: transmembrane #status predicted <TM8>  
F;352-368/Domain: transmembrane #status predicted <TM9>

Query Match 66.7%; Score 30; DB 2; Length 382;



Best Local Similarity 60.0%; Pred. NO. 83; Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:||| |||  
Db 277 LGSIAMLSQA 286

RESULT 20

B72707  
hypothetical protein F25024.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
C:Accession: T08946  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16519  
A:Accession: T08946  
A:Molecule type: DNA  
A:Residues: 1-425 <BEV>  
A:Cross-references: EMBL:AL078469; GSPDB:GN00062; ATSP:F25024.30  
A:Experimental source: cultivar Columbia; BAC clone F25024  
C:Genetics:  
A:Gene: ATSP:F25024.30  
A:Map position: 4  
A:Introns: 298/1

Query Match 66.7%; Score 30; DB 2; Length 425;  
Best Local Similarity 77.8%; Pred. NO. 93; Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIIVQLSQS 10  
:||| |||  
Db 279 GGIIVALSQS 287

RESULT 21

E82699  
carboxypeptidase related protein XFL1282 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: E82699  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82699  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-503 <SIM>  
A:Cross-references: GB:AE003962; GB:AE003849; NID:g9106270; PIDN:AAF84091.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tsuchako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; 2  
A:Reference number: A59328  
C:Genetics:  
A:Contents: annotation  
A:Gene: XFL1282

Query Match 66.7%; Score 30; DB 2; Length 503;

Best Local Similarity 66.7%; Pred. NO. 1.1e+02; Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
:||| |||  
Db 417 GSVTLRQS 425

RESULT 22

B72707  
hypothetical protein APE1073 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: B72707  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J. DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero A:Reference number: A72450; MUID:99310339  
A:Accession: B72707  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-539 <KAW>  
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA80058.1; PID:d1043844; PID:g A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1073

Query Match 66.7%; Score 30; DB 2; Length 539;  
Best Local Similarity 70.0%; Pred. NO. 1.2e+02; Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:||| |||  
Db 396 MKEIVQLSES 405

RESULT 23

S30288  
C4-dicarboxylate sensor kinase - Rhodobacter capsulatus  
C:Species: Rhodobacter capsulatus  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Oct-1999  
C:Accession: S30288  
R:Hamblin, M.J.; Shaw, J.G.; Kelly, D.J. Mol. Gen. Genet. 237, 215-224, 1993  
A:Title: Sequence analysis and interposon mutagenesis of a sensor-kinase (DctS) and r er capsulatus.  
A:Reference number: S30288; MUID:93204897  
A:Accession: S30288  
A:Molecule type: DNA  
A:Residues: 1-657 <HAM>  
A:Cross-references: EMBL:X64733; NID:g287816; PIDN:CAA45999.1; PID:e45702; PID:g58149 A:Note: the authors translated the initiation codon CTG for residue 1 as Leu  
C:Genetics:  
A:Gene: dctS  
A:Start codon: CTG  
C:Superfamily: sensor histidine kinase homology  
C:Keywords: autophosphorylation; membrane protein; phosphohistidine; phosphoprotein F:408-649/domain: sensor histidine kinase homology <SHK>  
F:440/Binding site: phosphate (His) (covalent) (by autophosphorylation) #status predi  
Query Match 66.7%; Score 30; DB 2; Length 657;  
Best Local Similarity 40.0%; Pred. NO. 1.5e+02; Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:||| |||  
Db 399 MGSVIDITQA 408

RESULT 24

Tl17339  
hypothetical protein DKFZp4340225.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: Tl17339  
R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999  
A;Reference number: 218727  
A;Accession: Tl17339  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-662 <DUE>  
A;Cross-references: EMBL:AL117647  
A;Experimental source: adult testis; clone DKFZp4340225  
C;Genetics:  
A;Note: DKFZp4340225.1

Query Match 66.7%; Score 30; DB 2; Length 662;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVQLS 8  
||:||||  
Db 590 GSVVQLS 596

RESULT 25  
S75636  
sensory transduction histidine kinase slr1969 - Synechocystis sp. (strain PCC 6803)  
N;Alternate names: protein slr1969  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C;Accession: S75636  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.  
A;Reference number: S74322; MUID:97061201  
A;Accession: S75636  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-750 <KAN>  
A;Cross-references: EMBL:D90912; GB:AB001339; NID:gl653228; PIDN:BAA18197.1; PID:d101893  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Superfamily: response regulator homology  
C;Keywords: phosphoprotein  
F;632-741/Domain: response regulator homology <RRH>  
F;680/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 66.7%; Score 30; DB 2; Length 750;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
: ||:||||  
Db 437 INSIIQLSQ 445

RESULT 26  
B33926  
DNA-directed RNA polymerase (EC 2.7.7.6) chain A [validated] - Sulfolobus acidocaldarius  
C;Species: Sulfolobus acidocaldarius  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Aug-2000  
C;Accession: B33926; S04717  
R;Puehler, G.; Leffers, H.; Gropp, F.; Palm, P.; Klenk, H.P.; Lottspeich, F.; Garrett, R.  
Proc. Natl. Acad. Sci. U.S.A. 86, 4569-4573, 1989  
A;Title: Archaeobacterial DNA-dependent RNA polymerases testify to the evolution of the e  
A;Reference number: A33926; MUID:89282812  
A;Accession: B33926

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual  
A;Molecule type: DNA  
A;Residues: 1-880 <PUE>  
R;Puehler, G.; Lottspeich, F.; Zillig, W.  
Nucleic Acids Res. 17, 4517-4534, 1989  
A;Title: Organization and nucleotide sequence of the genes encoding the large subunit  
A;Reference number: S04714; MUID:89315197  
A;Accession: S04717  
A;Molecule type: DNA  
A;Residues: 1-311,'N',313-560,'N',562-610,'M',612-640,'M',642-880 <PU2>  
A;Cross-references: EMBL:X14818; NID:g46667; PIDN:CAA32925.1; PID:g46670  
C;Genetics:  
A;Gene: rpoA  
C;Function:  
A;Description: EC 2.7.7.6 [validated; MUID:89315197]  
C;Superfamily: Halobacterium DNA-directed RNA polymerase chain A  
C;Keywords: nucleotidyltransferase; transcription

Query Match 66.7%; Score 30; DB 1; Length 880;  
Best Local Similarity 77.8%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
||||:|||  
Db 170 GSIVKLSPS 178

RESULT 27  
A84693  
hypothetical protein At2g29140 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cross)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: A84693  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487  
A;Accession: A84693  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-964 <STO>  
A;Cross-references: GB:AE002093; NID:g3980417; PIDN:AAC95220.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g29140  
A;Map position: 2

Query Match 66.7%; Score 30; DB 2; Length 964;  
Best Local Similarity 75.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 9  
||||:|||  
Db 846 GKIVQMSQ 853

RESULT 28  
F84693  
hypothetical protein At2g29200 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cross)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: F84693  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487  
A;Accession: F84693  
A;Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-968 <STO>  
A:Cross-references: GB:AE002093; NID:g3980412; PIDN:AAC95215.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g29200  
A:Map position: 2

Query Match 66.7%; Score 30; DB 2; Length 968;  
Best Local Similarity 75.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 9  
Db 850 GKIVQMSQ 857

RESULT 29  
E84693  
hypothetical protein At2g29190 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84693  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: E84693  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-972 <STO>  
A:Cross-references: GB:AE002093; NID:g3980413; PIDN:AAC95216.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g29190  
A:Map position: 2

Query Match 66.7%; Score 30; DB 2; Length 972;  
Best Local Similarity 75.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 9  
Db 854 GKIVQMSQ 861

RESULT 30  
T12195  
sucrose-phosphate synthase (EC 2.4.1.14) - fava bean  
C:Species: Vicia faba (fava bean)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T12195  
R:Heim, U.; Weber, H.; Wobus, U.  
Gene 178, 201-203, 1996  
A:Title: Cloning and characterization of a full-length cDNA encoding sucrose phosphate  
A:Reference number: Z17449; MUID:97080571  
A:Accession: T12195  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1059 <HEI>  
A:Cross-references: EMBL:Z56278; NID:g1022364; PIDN:CAA91217.1; PID:g1022365  
A:Experimental source: cultivar Fribo, seed coat  
C:Genetics:  
A:Gene: SPS  
A:Function:  
A:Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose  
A:Pathway: sucrose biosynthesis  
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology  
C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis  
F:169-654/Domain: sucrose/sucrose-phosphate synthase homology <SSS>

Query Match 66.7%; Score 30; DB 2; Length 1059;  
Best Local Similarity 50.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
Db 286 MGHIIQMSKA 295

RESULT 31  
S53043  
probable membrane protein YMR012w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YMR012w.16  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 06-Feb-1998  
C:Accession: S53043  
R:Devlin, K.; Churcher, C.M.  
submitted to the EMBL Data Library, March 1995  
A:Reference number: S53028  
A:Accession: S53043  
A:Molecule type: DNA  
A:Residues: 1-1277 <DEV>  
A:Cross-references: EMBL:Z48613; NID:g728645; PID:g728661; MIPS:YMR012w  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: SGD:CLU1  
A:Cross-references: SGD:S0004614; MIPS:YMR012w  
A:Map position: 13R  
C:Keywords: transmembrane protein  
F:835-851/Domain: transmembrane #status predicted <TMM>

Query Match 66.7%; Score 30; DB 2; Length 1277;  
Best Local Similarity 55.6%; Pred. No. 2.9e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
Db 710 LGKIIELSQ 718

RESULT 32  
T17285  
hypothetical protein DKFZp434N0535.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T17285  
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999  
A:Reference number: Z18723  
A:Accession: T17285  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1341 <POU>  
A:Cross-references: EMBL:AL117518  
A:Experimental source: adult testis; clone DKFZp434N0535  
C:Genetics:  
A:Note: DKFZp434N0535.1

Query Match 66.7%; Score 30; DB 2; Length 1341;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVQLS 8  
Db 1269 GSVVQLS 1275

RESULT 33  
T23620  
hypothetical protein KL2D12.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C;Accession: T23620; T28109  
R;Coles, L.  
submitted to the EMBL Data Library, April 1995  
A;Reference number: Z19772  
A;Accession: T23620  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1520 <WIL>  
A;Cross-references: EMBL:249069; PIDN:CAA88867.1; GSPDB:GN00020; CESP:K12D12.1  
R;Swinburne, J.  
submitted to the EMBL Data Library, March 1996  
A;Reference number: Z20470  
A;Accession: T28109  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1520 <WI2>  
A;Cross-references: EMBL:270213; PIDN:CAA94177.1; GSPDB:GN00020; CESP:K12D12.1  
A;Experimental source: clone ZK930  
C;Genetics:  
A;Gene: CESP:K12D12.1  
A;Map position: 2  
A;Introns: 34/1; 146/2; 390/3; 471/3; 611/2; 1351/3; 1486/2  
C;Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd

Query Match 66.7%; Score 30; DB 2; Length 1520;  
Best Local Similarity 66.7%; Pred. No. 3.5e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
|||:|:|  
Db 800 MGTIVNLAQ 808

RESULT 34  
T09144  
Probable guanine nucleotide exchange factor RhoGEF2 - fruit fly (Drosophila melanogaster)  
N;Alternate names: Shar pei/DRhoGEF2  
C;Species: Drosophila melanogaster  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 17-Nov-2000  
C;Accession: T09144; T09223  
R;Haecker, U.; Perrimon, N.  
submitted to the EMBL Data Library, October 1997  
A;Reference number: Z16586  
A;Accession: T09144  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-2559 <HAE>  
A;Cross-references: EMBL:AF031930; NID:g2687355; PID:g2687356  
R;Barrett, K.; Leptin, M.; Settleman, J.  
Cell 91, 905-915, 1997  
A;Title: The Rho GTPase and a putative RhoGEF mediate a signaling pathway for the cell s  
A;Reference number: Z16618; MUID:98088790  
A;Accession: T09223  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-676,'L',678-837,'L',839-889,'S',891-1243,'D',1245-1358,'E',1360-1368,'R',1  
A;Cross-references: EMBL:AF032870; NID:g2760367; PIDN:AAC38820.1; PID:g2760368  
C;Genetics:  
A;Gene: rhoGEF2  
A;Cross-references: FlyBase:FBgn0023172  
A;Map position: 2; 53F1-2  
A;Note: orchestrates cell shape changes during gastrulation  
C;Function:  
A;Description: mediates actin rearrangements required for cell shape changes during gastr  
C;Superfamily: protein kinase C zinc-binding repeat homology  
C;Keywords: signal transduction; embryo; GTP exchange  
F;1151-1200/Domain: protein kinase C zinc-binding repeat homology <KZ2>

Query Match 66.7%; Score 30; DB 2; Length 2559;  
Best Local Similarity 70.0%; Pred. No. 5.9e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
|||:|:|  
Db 196 MGGVQLNQS 205

RESULT 35  
T17464  
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei  
C;Species: Amycolatopsis mediterranei  
C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 17-Nov-2000  
C;Accession: T17464  
R;Schupp, T.  
submitted to the EMBL Data Library, December 1997  
A;Reference number: Z18802  
A;Accession: T17464  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-5069 <SCH>  
A;Cross-references: EMBL:AJ223012; NID:el227119; PID:el227121; PIDN:CAA11036.1  
A;Experimental source: strain LBG A3136  
C;Superfamily: acyl carrier protein homology  
C;Keywords: carrier protein  
F;1631-1702/Domain: acyl carrier protein homology <ACP1>  
F;3238-3309/Domain: acyl carrier protein homology <ACP2>  
F;4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 66.7%; Score 30; DB 2; Length 5069;  
Best Local Similarity 50.0%; Pred. No. 1.2e+03;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
|||:|:|  
Db 4219 LGAVVELPQS 4228

RESULT 36  
S12393  
hypothetical protein (qacA 5' region) - Staphylococcus aureus plasmid pSK1  
C;Species: Staphylococcus aureus  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Oct-1999  
C;Accession: S12393  
R;Rouch, D.A.; Cram, D.S.; DiBerardino, D.; Littlejohn, T.G.; Skurray, R.A.  
Mol. Microbiol. 4, 2051-2062, 1990  
A;Title: Efflux-mediated antiseptic resistance gene qacA from Staphylococcus aureus:  
A;Reference number: S12393; MUID:91211614  
A;Accession: S12393  
A;Molecule type: DNA  
A;Residues: 1-188 <ROU>  
A;Cross-references: EMBL:X56628; NID:g773395; PIDN:CAA39962.1; PID:g46660  
C;Genetics:  
A;Genome: plasmid

Query Match 64.4%; Score 29; DB 2; Length 188;  
Best Local Similarity 66.7%; Pred. No. 67;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
|||:|:|  
Db 26 GEIVKLSES 34

RESULT 37  
G86748  
hypothetical protein ykcb [imported] - Lactococcus lactis subsp. lactis (strain IL140  
C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 23-Mar-2001  
C;Accession: G86748



R: Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich Genome Res. in press, 2001  
 A: Title: The complete genome sequence of the lactic acid bacterium.  
 A: Reference number: A86625  
 A: Accession: G86748  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-225 <STO>  
 A: Cross-references: GB:AE005176; NID:gl2723935; PIDN:AAK05089.1; GSPDB:GN00146  
 A: Experimental source: strain IL1403  
 C: Genetics:  
 A: Gene: ykcb

Query Match 64.4%; Score 29; DB 2; Length 225;  
 Best Local Similarity 75.0%; Pred. No. 81;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIIVQLSQ 9  
 |||:||||  
 Db 198 GSIIQKSO 205

RESULT 38  
 F84145  
 transposase (23) BH3966 [imported] - Bacillus halodurans (strain C-125)  
 C: Species: Bacillus halodurans  
 C: Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
 C: Accession: F84145  
 R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000  
 A: Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A: Reference number: A83650; MUID:20263314  
 A: Accession: F84145  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-254 <STO>  
 A: Cross-references: GB:AP001520; GB:BA0000004; NID:gl0176401; PIDN:BA07685.1; GSPDB:GN00  
 A: Experimental source: strain C-125  
 C: Genetics:  
 A: Gene: BH3966  
 C: Superfamily: DNA replication protein dnaC

Query Match 64.4%; Score 29; DB 2; Length 254;  
 Best Local Similarity 60.0%; Pred. No. 91;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 |||:||||:  
 Db 137 MGELVQLLKS 146

RESULT 39  
 HB4149  
 transposase (27) BH4000 [imported] - Bacillus halodurans (strain C-125)  
 C: Species: Bacillus halodurans  
 C: Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
 C: Accession: HB4149  
 R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira Nucleic Acids Res. 28, 4317-4331, 2000  
 A: Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A: Reference number: A83650; MUID:20263314  
 A: Accession: HB4149  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-261 <STO>  
 A: Cross-references: GB:AP001520; GB:BA0000004; NID:gl0176401; PIDN:BA07719.1; GSPDB:GN00  
 A: Experimental source: strain C-125  
 C: Genetics:  
 A: Gene: BH4000  
 C: Superfamily: DNA replication protein dnaC

Query Match 64.4%; Score 29; DB 2; Length 261;  
 Best Local Similarity 60.0%; Pred. No. 94;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 |||:||||:  
 Db 144 MGELVQLLKS 153

RESULT 40  
 T29574  
 hypothetical protein F30B5.7 - Caenorhabditis elegans  
 C: Species: Caenorhabditis elegans  
 C: Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C: Accession: T29574  
 R: Miller, N.; Bradshaw, H.  
 A: Description: The sequence of C. elegans cosmid F30B5.  
 A: Reference number: 220644  
 A: Accession: T29574  
 A: Status: preliminary; translated from GB/EMBL/DDBJ  
 A: Molecule type: DNA  
 A: Residues: 1-274 <MIL>  
 A: Cross-references: EMBL:U42437; PIDN:AAA83496.1; CESP:F30B5.7  
 C: Genetics:  
 A: Gene: CESP:F30B5.7  
 A: Introns: 33/3; 79/2; 156/3

Query Match 64.4%; Score 29; DB 2; Length 274;  
 Best Local Similarity 60.0%; Pred. No. 99;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
 |||:||||:  
 Db 177 MGSILQNSSET 186

RESULT 41  
 H71938  
 flagellar motor switch protein - Helicobacter pylori (strain J99)  
 C: Species: Helicobacter pylori  
 A: Variety: strain J99  
 C: Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
 C: Accession: H71938  
 R: Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F. Nature 397, 176-180, 1999  
 A: Title: Genomic sequence comparison of two unrelated isolates of the human gastric p  
 A: Reference number: A71800; MUID:99120557  
 A: Accession: H71938  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-285 <ARN>  
 A: Cross-references: GB:AE001473; GB:AE001439; NID:g4154910; PIDN:AAD05971.1; PID:g415  
 A: Experimental source: strain J99  
 C: Genetics:  
 A: Gene: jhp0394

Query Match 64.4%; Score 29; DB 2; Length 285;  
 Best Local Similarity 55.6%; Pred. No. 1e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
 |||:||||:  
 Db 231 IGSVVELDQ 239

RESULT 42  
 F64648  
 fly protein - Helicobacter pylori (strain 26695)

C;Species: Helicobacter pylori  
C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C;Accession: F64648  
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne-son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997  
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.; A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A;Reference number: A64520; MUID:97394467  
A;Accession: F64648  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-287 <TOM>  
A;Cross-references: GB:AE000611; GB:AE000511; NID:g2314173; PIDN:AAD08074.1; PID:g231417

Query Match 64.4%; Score 29; DB 2; Length 287;  
Best Local Similarity 55.6%; Pred. No. 1e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:||:|:| |  
Db 233 IGSVVQLDQ 241

RESULT 43  
A40573  
clathrin heavy chain - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 03-Apr-1992 #sequence\_revision 03-Apr-1992 #text\_change 13-Aug-1999  
C;Accession: A40573  
R;Dodge, G.R.; Kovalszky, I.; McBride, O.W.; Yi, H.F.; Chu, M.; Saitta, B.; Stokes, D.G.; Genomics 11, 174-178, 1991  
A;Title: Human clathrin heavy chain (CLTC): partial molecular cloning, expression, and m  
A;Reference number: A40573; MUID:92112210  
A;Accession: A40573  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-305 <DOD>  
A;Cross-references: GB:X55878; GB:S75467; NID:g29982; PIDN:CAA39363.1; PID:g29983  
C;Superfamily: clathrin heavy chain

Query Match 64.4%; Score 29; DB 2; Length 305;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:|||| |  
Db 156 LGSIVNFSQ 164

RESULT 44  
H70790  
hypothetical protein Rv3679 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: H70790  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987  
A;Accession: H70790  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-340 <COL>  
A;Cross-references: GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CAA18001.1; PID:el26453  
A;Experimental source: strain H37RV  
C;Genetics:

A;Gene: Rv3679

Query Match 64.4%; Score 29; DB 2; Length 340;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:||: ||| |  
Db 324 LGSLYELSES 333

RESULT 45  
A85877  
probable prophage DNA injection protein Z3614 [imported] - Escherichia coli (strain O  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: A85877  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May-iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda-Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: A85877  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-378 <STO>  
A;Cross-references: GB:AE005174; NID:gl2516716; PIDN:AAG57477.1; GSPDB:GN00145; UWGP:  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z3614

Query Match 64.4%; Score 29; DB 2; Length 378;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
:||: | || |  
Db 59 LGSVAOTSQA 68

Search completed: June 28, 2001, 11:57:06  
Job time: 247 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:29 ; Search time 22.61 seconds  
(without alignments)  
15.151 Million cell updates/sec

Title: US-09-439-313-573  
Perfect score: 45  
Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 65 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	68.9	304	1 PP13_TOBAC	O04858 nicotiana t
2	31	68.9	428	1 DCTA_SALTY	P50334 salmonella
3	31	68.9	468	1 PPAR_HUMAN	Q07869 homo sapien
4	31	68.9	468	1 PPAR_MOUSE	P23204 mus musculu
5	31	68.9	468	1 PPAR_RAT	P37230 rattus norv
6	31	68.9	490	1 C7DB_LOTJA	Q22307 lotus japon
7	31	68.9	838	1 CYAA_PASMU	Q05766 pasteurella
8	31	68.9	1508	1 AT5A_MOUSE	O54827 mus musculu
9	30	66.7	305	1 Y031_STRGR	P36261 streptomyce
10	30	66.7	316	1 PP17_ARATH	O82733 arabidopsis
11	30	66.7	382	1 YCAD_ECOLI	P21503 escherichia
12	30	66.7	657	1 DCTS_RHOCA	P37739 rhodobacter
13	30	66.7	880	1 RPA1_SULAC	P11512 sulfolobus
14	30	66.7	1059	1 SPS_VICFA	Q43876 vicia faba
15	30	66.7	1277	1 IF3X_YEAST	Q03690 saccharomyc
16	30	66.7	1520	1 TOP2_CAEEL	Q23670 caenorhabdi
17	29	64.4	188	1 YP23_STAAU	P23217 staphylococ
18	29	64.4	340	1 Y0H9_MYCTU	O69647 mycobacteri
19	29	64.4	384	1 NAPP_YEAST	P53164 saccharomyc
20	29	64.4	459	1 DHE4_EMENI	P18819 emericeella
21	29	64.4	463	1 SELA_ECOLI	P23328 escherichia
22	29	64.4	503	1 VGLY_P1ARV	P03540 pichinde ar
23	29	64.4	511	1 PUR1_DROME	Q27601 drosophila
24	29	64.4	540	1 KNLC_CAEEL	P46822 caenorhabdi
25	29	64.4	622	1 SKN7_YEAST	P38889 saccharomyc
26	29	64.4	1171	1 DIA3_MOUSE	Q9z207 mus musculu
27	29	64.4	1640	1 CLH2_HUMAN	P53675 homo sapien
28	29	64.4	1675	1 CLH1_HUMAN	Q00610 homo sapien
29	29	64.4	1675	1 CLH_BOVIN	P49951 bos taurus
30	29	64.4	1675	1 CLH_RAT	P11442 rattus norv
31	29	64.4	1681	1 CLH_CAEEL	P34574 caenorhabdi
32	29	64.4	2787	1 TEL1_YEAST	P38110 saccharomyc
33	28	62.2	337	1 G3P_MYCGE	P47543 mycoplasma

34	28	62.2	340	1	SED5_YEAST	Q01590 saccharomyc
35	28	62.2	387	1	SCN1_SCHPO	P41890 schizosacch
36	28	62.2	407	1	COAT_BBV	P04329 black beetl
37	28	62.2	428	1	DCTA_ECOLI	P37312 escherichia
38	28	62.2	496	1	SMVA_SALTY	P37594 salmonella
39	28	62.2	501	1	CPJ5_MOUSE	O54749 mus musculu
40	28	62.2	579	1	YHVO_YEAST	P38848 saccharomyc
41	28	62.2	614	1	YEHQ_ECOLI	P33353 escherichia
42	28	62.2	791	1	KDGL_DROME	Q01583 drosophila
43	28	62.2	910	1	PERT_BORPE	P14283 bordetella
44	28	62.2	1088	1	TYCA_BACBR	P09095 bacillus br
45	28	62.2	1178	1	MN4_YEAST	P36044 saccharomyc
46	28	62.2	1306	1	MSB2_YEAST	P32334 saccharomyc
47	28	62.2	1314	1	Y197_HUMAN	Q12769 homo sapien
48	28	62.2	2145	1	U520_CAEEL	Q9u2g0 caenorhabdi
49	28	62.2	2452	1	RPB1_PLAFD	P14248 plasmodium
50	27	60.0	108	1	SUI1_YEAST	P32911 saccharomyc
51	27	60.0	146	1	MGN_HUMAN	P50606 homo sapien
52	27	60.0	146	1	MGN_XENLA	O42149 xenopus lae
53	27	60.0	197	1	Y928_METJA	Q58338 methanococc
54	27	60.0	209	1	YC22_PORPU	P51372 porphyra pu
55	27	60.0	233	1	LPPX_MYCTU	P96286 mycobacteri
56	27	60.0	307	1	CD5R_BOVIN	Q28199 bos taurus
57	27	60.0	307	1	CD5R_HUMAN	Q15078 homo sapien
58	27	60.0	307	1	CD5R_MOUSE	Q62938 mus musculu
59	27	60.0	311	1	PYRB_LACPL	P77883 lactobacill
60	27	60.0	320	1	ASPG_SPOFR	O02467 spodoptera
61	27	60.0	360	1	RTCA_AQUAE	O66884 aquifex aeo
62	27	60.0	365	1	RECA_SPIPL	P48293 spirulina p
63	27	60.0	366	1	VU1_HSV6U	Q01349 human herpe
64	27	60.0	408	1	GCST_ARATH	O65396 arabidopsis
65	27	60.0	447	1	SYH_SYNY3	Q55653 synechocyst

ALIGNMENTS

RESULT 1	PP13_TOBAC	STANDARD;	PRT;	304 AA.
ID	PP13_TOBAC			
AC	O04858;			
DT	01-OCT-2000 (Rel. 40, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	SERINE/THREONINE PROTEIN PHOSPHATASE PPI ISOZYME 3 (EC 3.1.3.16).			
GN	NPP3			
OS	Nicotiana tabacum (Common tobacco).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;			
OC	Solanales; Solanaceae; Nicotiana.			
OX	NCBI_TaxID=4097;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. XANTHI;			
RX	MEDLINE=98145437; PubMed=9484443;			
RA	Suh M., Cho H., Kim Y., Liu J., Lee H.;			
RT	"Multiple genes encoding serine/threonine protein phosphatases and			
RT	their differential expression in Nicotiana tabacum.";			
RL	Plant Mol. Biol. 36:315-322(1998).			
CC	-!- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +			
CC	ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).			
CC	-!- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-1			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; Z93770; CAB07805.1; -.			
DR				





```
CC HYPOLIPIDEMIC DRUGS AND FATTY ACIDS. ONCE ACTIVATED BY A LIGAND,
CC THE RECEPTOR BINDS TO A PROMOTER ELEMENT IN THE GENE FOR ACYL-COA
CC OXIDASE AND ACTIVATES ITS TRANSCRIPTION. IT THEREFORE CONTROLS THE
CC PEROXISOMAL BETA-OXIDATION PATHWAY OF FATTY ACIDS.
CC -!- SUBUNIT: HETERODIMER WITH THE RETINOID X RECEPTOR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L02932; AAA36468.1; -.
CC EMBL; Y07619; CAA68898.1; -.
CC EMBL; S74349; AAB32649.1; -.
CC EMBL; AL078611; CAB44427.1; -.
CC EMBL; A49289; A49289.
CC HSSP; P03372; 1HCQ.
CC MIM; 170998; -.
CC InterPro; IPR000536; -.
CC InterPro; IPR001628; -.
CC InterPro; IPR001723; -.
CC InterPro; IPR003074; -.
CC InterPro; IPR003076; -.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00047; STROIDFINGER.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR01288; PROXISOMEPAAR.
CC PRINTS; PR01289; PROXISOMPAAR.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; Activator; DNA-binding;
KW Nuclear protein; Zinc-finger; Multigene family.
FT DNA_BIND 102 166 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 102 122 C4-TYPE.
FT ZN_FING 139 161 C4-TYPE.
FT DOMAIN 281 468 LIGAND-BINDING (POTENTIAL).
FT CONFLICT 71 71 T -> M (IN REF. 2).
FT CONFLICT 123 123 K -> M (IN REF. 2).
FT CONFLICT 268 268 A -> V (IN REF. 1).
FT CONFLICT 296 296 G -> A (IN REF. 1).
FT CONFLICT 444 444 V -> A (IN REF. 2).
SQ SEQUENCE 468 AA; 52225 MW; 850846FD51ADA883 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 468;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 31 MGNIQEISQS 40
||:|:|
RESULT 4
PPAR_MOUSE
ID PPAR_MOUSE STANDARD; PRT; 468 AA.
AC P23204;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (PPAR-ALPHA).
GN PPARA OR NR1C1 OR PPAR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=91015382; PubMed=2129546;
RA Issemann I., Green S.;
RT "Activation of a member of the steroid hormone receptor superfamily
RT by peroxisome proliferators.";
RL Nature 347:645-650(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=94168583; PubMed=8123021;
RA Gearing K.L., Crickmore A., Gustafsson J.-A.;
RT "Structure of the mouse peroxisome proliferator activated receptor
RT alpha gene.";
RL Biochem. Biophys. Res. Commun. 199:255-263(1994).
RN [3]
RP SEQUENCE OF 413-468 FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Liver;
RX MEDLINE=96061953; PubMed=7588749;
RA Jones P.S., Savory R., Barratt P., Bell A.R., Gray T.J.B.,
RA Jenkins N.A., Gilbert D.J., Copeland N.G., Bell D.R.;
RT "Chromosomal localisation, inducibility, tissue-specific expression
RT and strain differences in three murine
RT peroxisome-proliferator-activated-receptor genes.";
RL Eur. J. Biochem. 233:219-226(1995).
CC -!- FUNCTION: RECEPTOR THAT BIND PEROXISOME PROLIFERATORS SUCH AS
CC HYPOLIPIDEMIC DRUGS AND FATTY ACIDS. ONCE ACTIVATED BY A LIGAND,
CC THE RECEPTOR BINDS TO A PROMOTER ELEMENT IN THE GENE FOR ACYL-COA
CC OXIDASE AND ACTIVATES ITS TRANSCRIPTION. IT THEREFORE CONTROLS THE
CC PEROXISOMAL BETA-OXIDATION PATHWAY OF FATTY ACIDS.
CC -!- SUBUNIT: HETERODIMER WITH THE RETINOID X RECEPTOR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER, KIDNEY AND
CC HEART. VERY WEAKLY EXPRESSED IN BRAIN AND TESTIS.
CC -!- DEVELOPMENTAL STAGE: IT APPEARS FIRST AT DAY 13.5 POSTCONCEPTION,
CC AND INCREASES UNTIL BIRTH.
CC -!- DISEASE: PEROXISOME PROLIFERATORS ARE A DIVERSE GROUP OF
CC CHEMICALS THAT INCLUDE HYPOLIPIDAEMIC DRUGS, HERBICIDES AND
CC INDUSTRIAL PLASTICISERS. ADMINISTRATION OF THESE CHEMICALS TO
CC RODENTS RESULTS IN THE DRAMATIC PROLIFERATION OF HEPATIC
CC PEROXISOMES AS WELL AS LIVER HYPERPLASIA.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X57638; CAA40856.1; -.
CC EMBL; X75289; CAA53042.1; -.
CC EMBL; X75290; CAA53042.1; JOINED.
CC EMBL; X75291; CAA53042.1; JOINED.
CC EMBL; X75292; CAA53042.1; JOINED.
CC EMBL; X75293; CAA53042.1; JOINED.
CC EMBL; X75294; CAA53042.1; JOINED.
CC EMBL; X89577; CAA61754.1; -.
CC PIR; S11659; S11659.
CC HSSP; P03372; 1HCQ.
CC TRANSFAC; T00694; -.
CC MGD; MGI:104740; Ppara.
CC InterPro; IPR000536; -.
CC InterPro; IPR001628; -.
CC InterPro; IPR001723; -.
CC InterPro; IPR003074; -.
CC InterPro; IPR003076; -.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00047; STROIDFINGER.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR01288; PROXISOMEPAAR.
```

DR PRINTS; PRO1289; PROXISOMPAAR.  
DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor; Transcription regulation; Activator; DNA-binding;  
KW Nuclear protein; Zinc-finger; Multigene family.  
FT DNA\_BIND 102 166 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 102 122 C4-TYPE.  
FT ZN\_FING 139 161 C4-TYPE.  
FT DOMAIN 281 468 LIGAND-BINDING (POTENTIAL).  
FT CONFLICT 75 75 A -> R (IN REF. 1).  
SQ SEQUENCE 468 AA; 52347 MW; 2930A5191C610B6B CRC64;

Query Match 68.9%; Score 31; DB 1; Length 468;  
Best Local Similarity 60.0%; Pred. No. 19;  
Matches 6; Conservative 3; Mismatches 1; Indels 1; Gaps 0;

QY 1 MGSIVQLSQS 10  
DB 31 MGNIQEIQS 40

RESULT 5

PPAR\_RAT  
ID PPAT\_RAT STANDARD; PRT; 468 AA.  
AC P37230;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (PPAR-ALPHA).  
GN PPARA OR NR1C1 OR PPAR.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-92262498; PubMed-1316614;  
RA Goettlicher M., Widmark E., Li Q., Gustafsson J.-A.;  
RT "Fatty acids activate a chimera of the clofibrate acid-activated  
receptor and the glucocorticoid receptor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:4653-4657(1992).  
CC -1- FUNCTION: RECEPTOR THAT BIND PEROXISOME PROLIFERATORS SUCH AS  
HYPO-LIPIDEMIC DRUGS AND FATTY ACIDS. ONCE ACTIVATED BY A LIGAND,  
THE RECEPTOR BINDS TO A PROMOTER ELEMENT IN THE GENE FOR ACYL-COA  
OXIDASE AND ACTIVATES ITS TRANSCRIPTION. IT THEREFORE CONTROLS THE  
PEROXISOMAL BETA-OXIDATION PATHWAY OF FATTY ACIDS.  
CC -1- SUBUNIT: HETERODIMER WITH THE RETINOID X RECEPTOR.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
CC NR1 SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M88592; AAA41918.1; -  
DR PIR; A45288; A45288.  
DR HSSP; P03372; LHCO.  
DR InterPro; IPR000536; -  
DR InterPro; IPR001628; -  
DR InterPro; IPR001723; -  
DR InterPro; IPR003074; -  
DR InterPro; IPR003076; -  
DR Pfam; PF00104; hormone\_rec; 1.  
DR Pfam; PF00105; zf-C4; 1.  
DR PRINTS; PR00047; STROIDFINGER.  
DR PRINTS; PR00398; STRDHORMONER.  
DR PRINTS; PR01288; PROXISOMEPAR.  
DR PRINTS; PR01289; PROXISOMPAAR.

DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor; Transcription regulation; Activator; DNA-binding;  
KW Nuclear protein; Zinc-finger; Multigene family.  
FT DNA\_BIND 102 166 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 102 122 C4-TYPE.  
FT ZN\_FING 139 161 C4-TYPE.  
FT DOMAIN 281 468 LIGAND-BINDING (POTENTIAL).  
FT CONFLICT 75 75 A -> R (IN REF. 1).  
SQ SEQUENCE 468 AA; 52377 MW; 2A89E7D715C8DBA9 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 468;  
Best Local Similarity 60.0%; Pred. No. 19;  
Matches 6; Conservative 3; Mismatches 1; Indels 1; Gaps 0;

QY 1 MGSIVQLSQS 10  
DB 31 MGNIQEIQS 40

RESULT 6

C7DB\_LOTJA  
ID C7DB\_LOTJA STANDARD; PRT; 490 AA.  
AC O22307;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 71D11 (EC 1.14.-.-) (FRAGMENT).  
GN CYP71D11.  
OS Lotus japonicus.  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;  
OC Fabales; Fabaceae; Papilionoideae; Lotus.  
OX NCBI\_TaxID=34305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-CV. GIFU; TISSUE-Root nodules;  
RX MEDLINE-97422886; PubMed-9276951;  
RA Szczygłowski K., Hamburger D., Kapranov P., de Bruijn F.J.;  
RT "Construction of a Lotus japonicus late nodulin expressed sequence  
tag library and identification of novel nodule-specific genes.";  
RL Plant Physiol. 114:1335-1346(1997).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF000403; AAB69644.1; -  
DR InterPro; IPR001128; -  
DR Pfam; PF00067; P450; 1.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Heme.  
FT NON\_TER 1  
FT BINDING 427 427 HEME (BY SIMILARITY).  
SQ SEQUENCE 490 AA; 56023 MW; B6C83F2129D58247 CRC64;

Query Match 68.9%; Score 31; DB 1; Length 490;  
Best Local Similarity 66.7%; Pred. No. 20;  
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

QY 2 GSVQLSQS 10  
DB 152 GSVVNLQA 160

RESULT 7

CYAA\_PASMU  
ID CYAA\_PASMU STANDARD; PRT; 838 AA.

AC Q05766;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE) (ADENYLYL  
DE CYCLASE).  
GN CYA.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CNP1 / NTCC 10322;  
RX MEDLINE=92011391; PubMed=1917858;  
RA Mock M., Crasnier M., Duflet E., Dumay V., Danchin A.;  
RT "Structural and functional relationships between Pasteurella  
multocida and enterobacterial adenylate cyclases.";  
RL J. Bacteriol. 173:6265-6269(1991).  
RN [2]  
RP REVIEW.  
RX MEDLINE=93119764; PubMed=8418825;  
RA Danchin A.;  
RT "Phylogeny of adenylate cyclases.";  
RL Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).  
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M68901; AAA25532.1; -.  
DR PIR; A38172; A38172.  
DR InterPro; IPR000274; -.  
DR Pfam; PF01295; Adenylate\_cyclase\_1;  
DR PROSITE; PS01092; ADENYLATE\_CYCLASE\_1\_1; 1.  
DR PROSITE; PS01093; ADENYLATE\_CYCLASE\_1\_2; 1.  
KW Lyase; CAMP synthesis.  
FT DOMAIN 1 541 CATALYTIC (POTENTIAL).  
FT DOMAIN 547 838 REGULATORY (POTENTIAL).  
SQ SEQUENCE 838 AA; 96816 MW; 2DDFB43AC08C29C6 CRC64;  
  
Query Match 68.9%; Score 31; DB 1; Length 838;  
Best Local Similarity 60.0%; Pred. No. 37;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 MGSIVQLSQS 10  
IIII :II:  
Db 104 MGSIASISQT 113  
  
RESULT 8  
AT5A\_MOUSE STANDARD; PRT; 1508 AA.  
AC O54827;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE VA (EC 3.6.1.-).  
GN ATPC5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Teratocarcinoma;  
RX MEDLINE=20473714; PubMed=11015572;  
RA Halleck M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,  
RA Hacker C., Pyle S., Newman J.T., Nakanishi Y., Ando H., Weinstein D.,  
RA Williamson P., Schlegel R.A.;  
RT "Differential expression of putative transbilayer amphipath  
RT transporters.";  
RL Physiol. Genomics 1:139-150(1999).  
RN [2]  
RP SEQUENCE OF 16-435 FROM N.A.  
RX MEDLINE=98217376; PubMed=9548971;  
RA Halleck M.S., Pradhan D., Blackman C.F., Berkes C., Williamson P.L.,  
RA Schlegel R.A.;  
RT "Multiple members of a third subfamily of P-type ATPases identified by  
RT genomic sequences and ESTs.";  
RL Genome Res. 8:354-361(1998).  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- TISSUE SPECIFICITY: FOUND IN TESTIS. ALSO DETECTED IN FETAL  
CC TISSUES.  
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2  
CC ATPASES). SUBFAMILY IV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF011337; AAC02902.1; -.  
DR EMBL; AF156549; AAF09447.1; -.  
DR MGD; MGI:1330809; Atpc5.  
DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
KW Multigene family.  
FT DOMAIN 1 79 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 80 101 POTENTIAL.  
FT DOMAIN 102 107 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 108 129 POTENTIAL.  
FT DOMAIN 130 313 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 314 335 POTENTIAL.  
FT DOMAIN 336 366 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 367 388 POTENTIAL.  
FT DOMAIN 389 1101 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1102 1122 POTENTIAL.  
FT DOMAIN 1123 1134 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1135 1154 POTENTIAL.  
FT DOMAIN 1155 1184 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1185 1206 POTENTIAL.  
FT DOMAIN 1207 1213 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1214 1236 POTENTIAL.  
FT DOMAIN 1237 1242 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1243 1263 POTENTIAL.  
FT DOMAIN 1264 1281 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1282 1306 POTENTIAL.  
FT DOMAIN 1307 1508 CYTOPLASMIC (POTENTIAL).  
FT MOD\_RES 431 431 PHOSPHORYLATION (BY SIMILARITY).  
FT METAL 1045 1045 MAGNESIUM (BY SIMILARITY).  
FT METAL 1049 1049 MAGNESIUM (BY SIMILARITY).  
FT DOMAIN 17 23 POLY-ARG.  
FT DOMAIN 471 474 POLY-GLU.  
FT CONFLICT 16 22 WRRPRRR -> KLAACK (IN REF. 2).  
FT CONFLICT 435 435 T -> L (IN REF. 2).  
SQ SEQUENCE 1508 AA; 168699 MW; DC9A0D99AF7EEB9E CRC64;  
  
Query Match 68.9%; Score 31; DB 1; Length 1508;  
Best Local Similarity 60.0%; Pred. No. 71;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MGSIVQLSQS 10



```
Db 1466 LGSVLQSGS 1475
-----
RESULT 9
ID Y031_STRGR STANDARD; PRT; 305 AA.
AC P36261;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 31.2 KDA PROTEIN IN RPLA-RPLJ INTERGENIC REGION (ORF31).
OS Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2-3-11;
RX MEDLINE=94314177; PubMed=8039667;
RA Kuberski S., Kasberg T., Distler J.;
RT "The nusG gene of Streptomyces griseus: cloning of the gene and
RL analysis of the A-factor binding properties of the gene product.";
FEMS Microbiol. Lett. 119:33-39(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X72787; CAA51299.1; -
CC PIR: S32237; S32237.
CC KW Hypothetical protein.
CC SEQUENCE 305 AA; 31280 MW; DCC7F06243C55F5A CRC64;
-----
Query Match 66.7%; Score 30; DB 1; Length 305;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 MGSIVQLSQS 10
DB 75 MGSVMSMKQS 84
-----
RESULT 10
ID P17_ARATH STANDARD; PRT; 316 AA.
AC O82733;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SERINE/THREONINE PROTEIN PHOSPHATASE PPI ISOZYME 7 (EC 3.1.3.16).
GN TOP7
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98278376; PubMed=9617814;
RA Lin Q., Li J., Smith R.D., Walker J.C.;
RT "Molecular cloning and chromosomal mapping of type one
RL serine/threonine protein phosphatases in Arabidopsis thaliana.";
plant Mol. Biol. 37:471-481(1998).
CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O -> A PROTEIN +
CC -1- ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, ROSETTES AND FLOWERS.
CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-1
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U80920; AAC39459.1; -
CC HSSP: P08129; IFJM.
CC InterPro: I3303; Arath; 1262; 33303.
CC InterPro: IPR000934; -
CC Pfam: PF00149; STPHPTase; 1.
CC PRINTS: PR00114; STPHPTase.
CC PROSITE: PS00125; SER_THR_PHOSPHATASE; 1.
KW Hydrolase; Iron; Manganese; Multigene family.
FT METAL 60
FT METAL 62
FT METAL 87
FT METAL 87
FT METAL 119
FT ACT_SITE 120
FT METAL 168
FT METAL 243
SQ SEQUENCE 316 AA; 36066 MW; 179B4E062B29E2B1 CRC64;
-----
Query Match 66.7%; Score 30; DB 1; Length 316;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 GSIVQLSQS 10
DB 21 GKIVQLSET 29
-----
RESULT 11
ID YCAD_ECOLI STANDARD; PRT; 382 AA.
AC P21503; P75834;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 41.4 KDA PROTEIN IN DMSC-PFLA INTERGENIC REGION (ORF Y).
GN YCAD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 12.7-28.0 min region on the linkage map.";
DNA Res. 3:137-155(1996).
```

RN [3]
RP SEQUENCE OF 1-265 FROM N.A.
RX STRAIN-K12 / C600; PubMed=3062312;
RM MEDLINE=89096500;
RA Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RT "Nucleotide sequence of the dmsABC operon encoding the anaerobic
RL dimethylsulphoxide reductase of Escherichia coli.";
RM Mol. Microbiol. 2:785-795(1988).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000192; AAC73984.1; -.
DR EMBL; D90727; BAA35630.1; -.
DR EMBL; D90728; BAA35633.1; -.
DR EMBL; J03412; AAA83847.1; -.
DR PIR; S03788; S03788.
DR EcoGene; EG11242; ycaD.
DR InterPro; IPR001066; -.
DR Pfam; PF00083; sugar\_tr; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 102 122 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT CONFLICT 265 265 L -> H (IN REF. 2).
SQ SEQUENCE 382 AA; 41431 MW; 9C0F47A928B7C949 CRC64;
Query Match 66.7%; Score 30; DB 1; Length 382;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MGSIVQLSQS 10
Db 277 LGSIAMLSQA 286
RESULT 12
DCTS\_RHOCA
ID DCTS\_RHOCA STANDARD; PRT; 657 AA.
AC P37739;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE C4-DICARBOXYLATE TRANSPORT SENSOR PROTEIN DCTS (EC 2.7.3.-).
GN DCTS.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI\_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-B10;
RM MEDLINE=93204897; PubMed=8455557;
RA Hamblin M.J., Shaw J.G., Kelly D.J.;
RT "Sequence analysis and interposon mutagenesis of a sensor-kinase
(R) and response-regulator (DctR) controlling synthesis of the
high-affinity C4-dicarboxylate transport system in Rhodobacter

capsulatus.";
RM Mol. Gen. Genet. 237:215-224(1993).
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR
CC INVOLVED IN THE TRANSPORT OF C4-DICARBOXYLATES. DCTS FUNCTIONS AS
CC A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT PHOSPHORYLATES DCTR IN
CC RESPONSE TO ENVIRONMENTAL SIGNALS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64733; CAA45999.1; -.
DR PIR; S30288; S30288.
DR InterPro; IPR000014; -.
DR InterPro; IPR000410; -.
DR InterPro; IPR001610; -.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 1.
DR Pfam; PF00512; signal; 1.
DR PRINTS; PR00344; BCTRLSENSOR.
KW Sensory transduction; Transferrase; Kinase; Phosphorylation;
KW Transmembrane; Inner membrane.
FT DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 27 51 POTENTIAL.
FT DOMAIN 52 252 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 253 273 POTENTIAL.
FT DOMAIN 274 657 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 407 422 INTER-DOMAIN LINKER (POTENTIAL).
FT MOD\_RES 440 440 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 657 AA; 70142 MW; 139D2CFC7CFFA69B CRC64;
Query Match 66.7%; Score 30; DB 1; Length 657;
Best Local Similarity 40.0%; Pred. No. 48;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 MGSIVQLSQS 10
Db 399 MGSVIDITQA 408
RESULT 13
RPAL\_SULAC
ID RPAL\_SULAC STANDARD; PRT; 880 AA.
AC P11512;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT A' (EC 2.7.7.6).
GN RPOA1 OR RPOA.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI\_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 33909 / NCIB 11770 / DSM 639;
RM MEDLINE=89315197; PubMed=2501756;
RA Puehler G., Lottspeich F., Zillig W.;
RT "Organization and nucleotide sequence of the genes encoding the large
RT subunits A, B and C of the DNA-dependent RNA polymerase of the
RT archaeobacterium Sulfolobus acidocaldarius.";
RL Nucleic Acids Res. 17:4517-4534(1989).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +



Query Match 66.7%; Score 30; DB 1; Length 1277;  
Best Local Similarity 55.6%; Pred. NO. 1e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:| |:| |  
Db 710 LGKIELSQ 718

RESULT 16  
TOP2\_CAEEEL STANDARD; PRT; 1520 AA.  
AC Q23670; Q27509;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PROBABLE DNA TOPOISOMERASE II (EC 5.99.1.3).  
GN K12D12.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Swinburne J.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Coles L.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT  
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II  
CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING  
CC OF DOUBLE-STRANDED DNA.  
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z70213; CAA94177.1; -.  
DR EMBL; Z49069; CAA94177.1; JOINED.  
DR EMBL; Z49069; CAA88867.1; -.  
DR EMBL; Z70213; CAA88867.1; JOINED.  
DR WormPep; K12D12.1; CE06184.

DR InterPro; IPR000947; -.  
DR InterPro; IPR001154; -.  
DR InterPro; IPR001241; -.  
DR InterPro; IPR002205; -.  
DR Pfam; PF00204; DNA\_topoisoiI; 1.  
DR Pfam; PF00521; DNA\_topoisoiV; 1.  
DR PRINTS; PR00418; TPI2FAMILY.  
DR PRINTS; PR00615; CCAATSUBUNTA.  
DR PRINTS; PR01158; TOPI2SMRASEII.

DR PROSITE; PS00177; TOPOISOMERASE\_II; 1.  
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.  
FT NP\_BIND 196 201 ATP (POTENTIAL).  
FT ACT\_SITE 840 840 DNA CLEAVAGE (BY SIMILARITY).  
SQ SEQUENCE 1520 AA; 172333 MW; 016226697F41C360 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 1520;

Best Local Similarity 66.7%; Pred. NO. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
| |:| | |  
Db 800 MGTIVNLAQ 808

RESULT 17  
YP23\_STAAU STANDARD; PRT; 188 AA.  
AC P23217;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN QACA 5'REGION (ORF 188).  
OS Staphylococcus aureus.  
OG Plasmid pSK1.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91211614; PubMed=2089219;  
RA Rouch D.A.; Cram D.S.; Diberardino D.; Littlejohn T.G.;  
RA Skurray R.A.;  
RT "Efflux-mediated antiseptic resistance gene qacA from Staphylococcus  
RT aureus: common ancestry with tetracycline- and sugar-transport  
RT proteins.";  
RL Mol. Microbiol. 4:2051-2062(1990).  
CC -!- FUNCTION: POSSIBLE TRANSCRIPTIONAL REPRESSOR OF THE QACA GENE.  
CC -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X56628; CAA39962.1; -.  
DR PIR; S12393; S12393.  
DR InterPro; IPR001647; -.  
DR Pfam; PF00440; tetr; 1.  
DR PRINTS; PR00455; HTHTETR.  
DR PROSITE; PS01081; HTH\_TETR\_FAMILY; 1.

KW Hypothetical protein; Transcription regulation; DNA-binding; Plasmid;  
KW Repressor.  
FT DNA\_BIND 24 43 H-T-H MOTIF (POTENTIAL).  
SQ SEQUENCE 188 AA; 22174 MW; 7B91E005C8D47322 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 188;  
Best Local Similarity 66.7%; Pred. NO. 20;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
| |:| | |  
Db 26 GEIVKLSES 34

RESULT 18  
YOH9\_MYCTU STANDARD; PRT; 340 AA.  
AC O69647;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE ATPASE RV3679.  
GN RV3679 OR MTV025.027.  
OS Mycobacterium tuberculosis.



OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RC Nature 393:537-544(1998).  
 CC -!- SIMILARITY: SOME, TO THE ARSA ATPASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL022121; CAAL8001.1; -  
 DR Tuberculin; Rv3679; -  
 KW Hypothetical protein; ATP-binding.  
 FT NP\_BIND 28 35 ATP (POTENTIAL).  
 SQ SEQUENCE 340 AA; 35856 MW; 64E42BDD95A7635C CRC64;  
 Query Match 64.4%; Score 29; DB 1; Length 340;  
 Best Local Similarity 50.0%; Pred. No. 39;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MGSIVQLSQS 10  
 Db :||:||||  
 324 LGSLYELSES 333  
 RESULT 19  
 ID NAPP\_YEAST STANDARD; PRT; 384 AA.  
 AC P53164;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE NADH PYROPHOSPHATASE (EC 3.6.1.-).  
 GN YGL067W  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=20334341; PubMed=10873676;  
 RA Xu W., Dunn C.A., Bessman M.J.;  
 RA "Cloning and characterization of the NADH pyrophosphatases from  
 RT Caenorhabditis elegans and Saccharomyces cerevisiae, members of a  
 RT Nudix hydrolase subfamily.";  
 RL Biochem. Biophys. Res. Commun. 273:753-758(2000).  
 CC -!- CATALYTIC ACTIVITY: NADH + H(2)O = AMP + NMNH.  
 CC -!- COFACTOR: REQUIRES DIVALENT IONS: MAGNESIUM OR MANGANESE.  
 CC -!- SUBUNIT: HOMODIMER.  
 CC -!- SIMILARITY: BELONGS TO THE NUDIX FAMILY. NUDC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z72589; CAA96771.1; -  
 DR SGD; S0003035; YGI067W.  
 DR InterPro; IPR000086; -  
 DR Pfam; PF00293; mutt; 1.  
 DR PRINTS; PR00502; MUTTDOMAIN.  
 DR PROSITE; PS00893; NUDIX; 1.  
 DR Hydrolase; NAD; Magnesium; Manganese.  
 KW DOMAIN 257 NUDIX BOX.  
 FT SEQUENCE 384 AA; 43516 MW; 96D6768CE2BC8F0B CRC64;  
 SQ SEQUENCE 384 AA; 43516 MW; 96D6768CE2BC8F0B CRC64;  
 Query Match 64.4%; Score 29; DB 1; Length 384;  
 Best Local Similarity 66.7%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 GSIVQLSQS 10  
 Db :||:||||  
 52 GDLVQLSNS 60  
 RESULT 20  
 ID DHE4\_EMENI STANDARD; PRT; 459 AA.  
 AC P18819;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE NADP-SPECIFIC GLUTAMATE DEHYDROGENASE (EC 1.4.1.4) (NADP-GDH).  
 GN GDHA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=5072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89384423; PubMed=2550758;  
 RA Hawkins A.R., Gurr S.J., Montague P., Kinghorn J.R.;  
 RA "Nucleotide sequence and regulation of expression of the Aspergillus  
 RT nidulans gdhA gene encoding NADp dependent glutamate dehydrogenase.";  
 RL Mol. Genet. 218:105-111(1989).  
 CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE + H(2)O + NADP(+) = 2-OXOGLUTARATE  
 CC + NH(3) + NADPH.  
 CC -!- SUBUNIT: HOMOHETEROMER.  
 CC -!- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X16121; CAA34252.1; -  
 DR PIR; S04904; S04904.  
 DR HSSP; P24295; 1AUP.  
 DR InterPro; IPR001625;  
 DR Pfam; PF00208; GLFV\_dehydrog; 1.  
 DR PRINTS; PR00082; GLFDHGRGNASE.  
 DR PROSITE; PS00074; GLFV\_DEHYDROGENASE; 1.  
 KW Oxidoreductase; NADP.  
 FT ACT\_SITE 114 114  
 SQ SEQUENCE 459 AA; 49608 MW; 682964399C002B7D CRC64;  
 BY SIMILARITY.

Query Match 64.4%; Score 29; DB 1; Length 459;  
Best Local Similarity 66.7%; Pred. No. 55;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10  
II:IIII  
Db 245 GSVVSLSDS 253

RESULT 21  
SELA\_ECOLI STANDARD; PRT; 463 AA.  
AC SELA\_ECOLI P78119;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE L-SERYL-TRNA(SEC) SELENIUM TRANSFERASE (EC 2.9.1.1) (CYSTEINYL-  
DE TRNA(SEC) SELENIUM TRANSFERASE) (SELENOCYSTEINE SYNTHASE)  
DE (SELENOCYSTEINYL-TRNA(SEC) SYNTHASE).  
GN SELA OR FDHA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91177883; PubMed=2007584;  
RA Forchhammer K., Leinfelder W., Boesmillier K., Veprek B., Boeck A.;  
RT "Selenocysteine synthase from Escherichia coli. Nucleotide sequence  
of the gene (sela) and purification of the protein.";  
RL J. Biol. Chem. 266:6318-6323(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=94316500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
region from 76.0 to 81.5 minutes.";  
RL Nucleic Acids Res. 22:2576-2586(1994).  
CC -!- FUNCTION: CONVERTS SERYL-TRNA-UCA TO SELENOCYSTEINYL-TRNA-UCA  
DURING SELENOPROTEIN BIOSYNTHESIS.  
CC -!- CATALYTIC ACTIVITY: L-SERYL-TRNA(SEC) + SELENOPHOSPHATE -  
L-SELENOCYSTEINYL-TRNA(SEC) + H(2)O + ORTHOPHOSPHATE.  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE. THIS REACTION REQUIRES ATP,  
MAGNESIUM AND REDUCED SELENIUM.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M64177; AAA24624.1; -.  
DR EMBL; U00039; AAB18568.1; -.  
DR EMBL; AE000436; AAC76615.1; -.  
DR PIR; A38730; A38730.  
DR ECO2DBASE; G050.7; 6TH EDITION.  
DR EcoGene; EG10941; sela.  
KW Transferase; Pyridoxal phosphate; Magnesium; Selenium.  
FT CONFLICT 7 F -> S (IN REF. 2).  
SQ SEQUENCE 463 AA; 50667 MW; A2AC08A84E8ECD01 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 463;  
Best Local Similarity 75.0%; Pred. No. 55;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 9  
II:IIII  
Db 262 GSLVDLSQ 269

RESULT 22  
VGLY\_P1ARV STANDARD; PRT; 503 AA.  
AC P03540;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE GLYCOPROTEIN POLYPROTEIN PRECURSOR [CONTAINS: GLYCOPROTEINS G1 AND  
DE G2].  
GN GPC.  
OS Pichinde arenavirus.  
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.  
OX NCBI\_TaxID=11630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85033957; PubMed=6492264;  
RA Auperin D.D., Romanowski V., Galinski M., Bishop D.H.L.;  
RT "Sequencing studies of pichinde arenavirus S RNA indicate a novel  
coding strategy, an ambisense viral S RNA.";  
RL J. Virol. 52:897-904(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87160943; PubMed=2435460;  
RA Bishop D.H.L., Auperin D.D.;  
RT "Arenavirus gene structure and organization.";  
RL Curr. Top. Microbiol. Immunol. 133:5-17(1987).  
CC -!- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; K02734; AAA46824.1; -.  
DR EMBL; M16735; AAA46827.1; -.  
DR PIR; A04149; QQXPGP.  
DR InterPro; IPR001535; -.  
DR Pfam; PF00798; Arena\_glycoprot; 1.  
KW Polyprotein; Glycoprotein; Envelope protein.  
FT CHAIN 1 271 GLYCOPROTEIN G1.  
FT CHAIN 272 503 GLYCOPROTEIN G2.  
FT CARBOHYD 67 67 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 111 111 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 132 132 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 241 241 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 379 379 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 503 AA; 57278 MW; 17740E092B450044 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 503;  
Best Local Similarity 70.0%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10  
II:IIII  
Db 1 MGQIVTLIQS 10





Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MGSIVQLSQS 10  
||| :|||

Db 497 MGGIDEMSQS 506

RESULT 25  
SKN7\_YEAST  
ID SKN7\_YEAST STANDARD; PRT; 622 AA.  
AC P38889; P39747;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PUTATIVE TRANSCRIPTION FACTOR SKN7 (POS9 PROTEIN).  
GN SKN7 OR POS9 OR BRY1 OR YHR206W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94042854; PubMed=8226633;  
RA Brown J.L., North S., Bussey H.;  
RT "SKN7, a yeast multicopy suppressor of a mutation affecting cell wall  
RT beta-glucan assembly, encodes a product with domains homologous to  
RT prokaryotic two-component regulators and to heat shock transcription  
RT factors.";  
RL J. Bacteriol. 175:6908-6915(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Krems B., Charizanis C., Entian K.-D.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,  
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
RA Vaudin M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
RT VIII.";  
RL Science 265:2077-2082(1994).  
RN [4]  
RP FUNCTION, AND MUTAGENESIS.  
RX MEDLINE=95045411; PubMed=7957083;  
RA Brown J.L., Bussey H., Stewart R.C.;  
RT "Yeast Skn7p functions in a eukaryotic two-component regulatory  
RT pathway.";  
RL EMBO J. 13:5186-5194(1994).  
CC -!- FUNCTION: INVOLVED IN OXIDATIVE STRESS. TRANSCRIPTION FACTOR THAT  
CC MAY FUNCTION IN A TWO-COMPONENT SIGNAL TRANSDUCTION PATHWAY THAT  
CC ACTS IN PARALLEL WITH THE PKC1 CASCADE TO REGULATE GROWTH AT THE  
CC CELL SURFACE.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.  
CC -!- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U00485; AAC48911.1; -;  
CC EMBL; X83031; CAA58143.1; -;  
CC EMBL; U00029; AAB69734.1; -;

PIR; A49344; A49344.  
PIR; S48987; S48987.  
DR HSSP; P22121; 2HTS.  
DR SGD; S0001249; SKN7.  
DR InterPro; IPR000232; -.  
DR InterPro; IPR001789; -.  
DR Pfam; PF00447; HSF\_DNA-bind; 1.  
DR Pfam; PF00072; response\_reg; 1.  
DR PRINTS; PR00056; HSFDOMAIN.  
DR PROSITE; PS00434; HSF\_DOMAIN; 1.  
KW Transcription regulation; Sensory transduction; Nuclear protein;  
KW DNA-binding; Phosphorylation.  
FT DNA\_BIND 86 190 BY SIMILARITY.  
FT MOD\_RES 427 427 PHOSPHORYLATION (PROBABLE).  
FT MUTAGEN 427 427 D->N: DIMINISHED ACTIVITY.  
FT MUTAGEN 427 427 D->E: AUGMENTED ACTIVITY.  
SQ SEQUENCE 622 AA; 69202 MW; 4C732FD66E326742 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 622;  
Best Local Similarity 70.0%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MGSIVQLSQS 10  
||| |||

Db 605 MGSTPQLPQS 614

RESULT 26  
DIA3\_MOUSE  
ID DIA3\_MOUSE STANDARD; PRT; 1171 AA.  
AC Q9Z207;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE DIAPHANOUS PROTEIN HOMOLOG 3 (DIAPHANOUS-RELATED FORMIN 3) (DRF3)  
DE (MDIA2) (P134MDIA2).  
GN DIAPH3 OR DIAP3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98204843; PubMed=9535835;  
RA Alberts A.S., Bouquin N., Johnston L.H., Treisman R.;  
RT "Analysis of RhoA-binding proteins reveals an interaction domain  
RT conserved in heterotrimeric G protein beta subunits and the yeast  
RT response regulator protein Skn7.";  
RL J. Biol. Chem. 273:8616-8622(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Tomimaga T., Sahai E., Treisman R.H., Alberts A.S.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP FUNCTION.  
RX MEDLINE=20142655; PubMed=10678165;  
RA Tomimaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,  
RA Alberts A.S.;  
RT "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase  
RT signaling.";  
RL Mol. Cell 5:13-25(2000).  
CC -!- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN  
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE  
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,  
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE  
CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE  
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS.  
CC -!- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE  
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION.  
CC -!- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).  
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.





```
QY      1 MGSIVQLSQ 9
      :||||| ||
Db      715 LGSIVNFSQ 723

RESULT 28
CLH1_HUMAN          STANDARD;          PRT; 1675 AA.
AC      Q00610;
DT      01-DEC-1992 (Rel. 24, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      CLATHRIN HEAVY CHAIN 1 (CLH-17).
GN      CLTC OR CLH17 OR KIAA0034.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Bone marrow;
RX      MEDLINE=96051387; PubMed=7584026;
RA      Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA      Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT      "Prediction of the coding sequences of unidentified human genes. I.
RT      The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT      analysis of randomly sampled cDNA clones from human immature myeloid
RT      cell line KG-1."
RL      DNA Res. 1:27-35(1994).
RN      [2]
RP      SEQUENCE OF 560-864 FROM N.A.
RC      TISSUE=Colon;
RX      MEDLINE=92112210; PubMed=1765375;
RA      Dodge G.R., Kovalszky I., McBride O.W., Yi H.F., Chu M.L., Saitta B.,
RA      Stokes D.G., Iozzo R.V.;
RT      "Human clathrin heavy chain (CLTC): partial molecular cloning,
RT      expression, and mapping of the gene to human chromosome 17q11-qter."
RL      Genomics 11:174-178(1991).
CC      -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC      COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
CC      LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
CC      TRANS GOLGI NETWORK.
CC      -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC      LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
CC      PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
CC      PH AND THE CONCENTRATION OF CALCIUM.
CC      -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC      VESICLES.
CC      -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; D21260; BAA04801.1; -.
DR      EMBL; X55878; CAA39363.1; -.
DR      PIR; A40573; A40573.
DR      MIM; 118955; -.
DR      InterPro; IPR000547; -.
DR      InterPro; IPR001473; -.
DR      Pfam; PF01394; Clathrin_propel; 7.
DR      Pfam; PF00637; Clathrin_repeat; 7.
KW      Coated pits.
FT      DOMAIN          1      479      GLOBULAR TERMINAL DOMAIN.
FT      DOMAIN          480      523      FLEXIBLE LINKER.
FT      DOMAIN          524      1675    HEAVY CHAIN ARM.
FT      DOMAIN          524      634      DISTAL SEGMENT.
FT      DOMAIN          639      1675    PROXIMAL SEGMENT.
FT      DOMAIN          449      465      BINDING SITE FOR THE UNCOATING ATPASE,
```

```
FT      INVOLVED IN LATTICE DISASSEMBLY
FT      (POTENTIAL).
FT      BINDING          1213      1522    LIGHT CHAIN (BY SIMILARITY).
FT      DOMAIN          1550      1675    TRIMERIZATION (BY SIMILARITY).
FT      CONFLICT          560      560      Q -> R (IN REF. 2).
FT      CONFLICT          817      817      G -> V (IN REF. 2).
SQ      SEQUENCE          1675 AA; 191614 MW; 6C4F2D54950079E2 CRC64;

Query Match          64.4%; Score 29; DB 1; Length 1675;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches          6; Conservative          1; Mismatches          2; Indels          0; Gaps          0;

QY      1 MGSIVQLSQ 9
      :||||| ||
Db      715 LGSIVNFSQ 723

RESULT 29
CLH_BOVIN
ID      CLH_BOVIN          STANDARD;          PRT; 1675 AA.
AC      P49951;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      CLATHRIN HEAVY CHAIN.
GN      CLTC.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Kidney;
RX      MEDLINE=96028100; PubMed=7585943;
RA      Liu S.-H., Wong M.L., Craik C.S., Brodsky F.M.;
RT      "Regulation of clathrin assembly and trimerization defined using
RT      recombinant triskelion hubs."
RL      Cell 83:257-267(1995).
CC      -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC      COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
CC      LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
CC      TRANS GOLGI NETWORK.
CC      -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC      LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
CC      PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
CC      PH AND THE CONCENTRATION OF CALCIUM.
CC      -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC      VESICLES.
CC      -!- DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF
CC      THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND
CC      THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE
CC      CLATHRIN LATTICE.
CC      -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; U31757; AAC48524.1; -.
DR      InterPro; IPR000547; -.
DR      InterPro; IPR001473; -.
DR      Pfam; PF01394; Clathrin_propel; 7.
DR      Pfam; PF00637; Clathrin_repeat; 7.
KW      Coated pits.
FT      DOMAIN          1      479      GLOBULAR TERMINAL DOMAIN.
FT      DOMAIN          480      523      FLEXIBLE LINKER.
FT      DOMAIN          524      1675    HEAVY CHAIN ARM.
```

FT DOMAIN 524 634 DISTAL SEGMENT.  
FT DOMAIN 639 1675 PROXIMAL SEGMENT.  
FT DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,  
INVOLVED IN LATTICE DISASSEMBLY  
(POTENTIAL).  
FT BINDING 1213 1522 LIGHT CHAIN.  
FT DOMAIN 1550 1675 TRIMERIZATION.  
FT SEQUENCE 1675 AA; 191587 MW; 6C4F2D54801579E2 CRC64;  
SQ  
Query Match 64.4%; Score 29; DB 1; Length 1675;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MGSIVQLSQ 9  
Db 715 LGSIVNFSQ 723  
RESULT 30  
CLH\_RAT STANDARD; PRT; 1675 AA.  
ID CLH\_RAT  
AC P11442; 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CLATHRIN HEAVY CHAIN.  
GN CLTC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=101116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88097376; PubMed=3480512;  
RA Kirchhausen T.; Harrison S.C.; Chow E.P.; Mattaliano R.J.;  
RA Ramachandran K.L.; Smart J.; Brosius J.;  
RT "Clathrin heavy chain: molecular cloning and complete primary  
structure.";  
RT Proc. Natl. Acad. Sci. U.S.A. 84:8805-8809(1987).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-493.  
RX MEDLINE=99043510; PubMed=9827808;  
RA Ter Haar E.; Musacchio A.; Harrison S.C.; Kirchhausen T.;  
RT "Atomic structure of clathrin: a beta propeller terminal domain joins  
an alpha zigzag linker.";  
RT Cell 95:563-573(1998).  
RN [1]  
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES  
LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE  
TRANS GOLGI NETWORK.  
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3  
LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE  
PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE  
PH AND THE CONCENTRATION OF CALCIUM.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
VESICLES.  
CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: J03583; AAA40874.1;  
CC PIR: A39941; LRRTH.  
CC PDB: 1BPO; 06-APR-99.  
CC InterPro: IPR000547;  
CC InterPro: IPR001473;  
-----

DR Pfam: PF01394; Clathrin\_propel; 7.  
DR Pfam: PF00637; Clathrin\_repeat; 7.  
KW Coated pits; 3D-structure.  
FT DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.  
FT DOMAIN 480 523 FLEXIBLE LINKER.  
FT DOMAIN 524 1675 HEAVY CHAIN ARM.  
FT DOMAIN 524 634 DISTAL SEGMENT.  
FT DOMAIN 639 1675 PROXIMAL SEGMENT.  
FT SEQUENCE 1675 AA; 191598 MW; C10F54C7ED8C5A61 CRC64;  
SQ  
Query Match 64.4%; Score 29; DB 1; Length 1675;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MGSIVQLSQ 9  
Db 715 LGSIVNFSQ 723  
RESULT 31  
CLH\_CAEEL STANDARD; PRT; 1681 AA.  
ID CLH\_CAEEL  
AC P34574;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PROBABLE CLATHRIN HEAVY CHAIN.  
GN T20G5.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Berk M.; Smith A.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
COATED PITS AND VESICLES (BY SIMILARITY).  
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3  
LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT  
(BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
VESICLES (BY SIMILARITY).  
CC -!- DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF  
THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND  
THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE  
CLATHRIN LATTICE.  
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: Z30423; CAA83003.1;  
CC PIR: S42369; S42369.  
CC WormPep: T20G5.1; CE00480.  
CC InterPro: IPR000547;  
CC InterPro: IPR001473;  
CC Pfam: PF01394; Clathrin\_propel; 7.  
CC Pfam: PF00637; Clathrin\_repeat; 7.  
KW Hypothetical protein; Coated pits.  
SQ SEQUENCE 1681 AA; 191541 MW; 44D15C61339009D9 CRC64;



Query Match 64.4%; Score 29; DB 1; Length 1681;  
Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
:|||||  
Db 717 LGSIVNFSQ 725

RESULT 32  
TELL\_YEAST STANDARD; PRT; 2787 AA.  
AC P38110;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE TELOMER LENGTH REGULATION PROTEIN TELL.  
GN TELL1 OR YBL088C OR YBL0706.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=96076635; PubMed=7502586;  
RA Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;  
RT "Sequence analysis of a 78.6 kb segment of the left end of  
Saccharomyces cerevisiae chromosome II.";  
RL Yeast 11:1103-1112(1995).  
RN [2]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RA Greenwell P.S., Kronmal S.L., Porter S.E., Gassenhuber J.,  
RA Obermaier B., Petes T.D.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PUTATIVE PHOSPHATIDYLINOSITOL KINASE INVOLVED IN  
CC CONTROLLING TELOMERE LENGTH.  
CC -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X79489; CAA56016.1; -  
CC EMBL; Z35849; CAA84909.1; -  
CC EMBL; U31331; AAA69802.1; -  
CC PIR; S45416; S45416.  
CC SGD; S0000184; TELL1.  
CC InterPro; IPR000403; -  
CC Pfam; PF00454; PI3\_PI4\_kinase; 1.  
CC PROSITE; PS00915; PI3\_4\_KINASE\_1; 1.  
CC PROSITE; PS00916; PI3\_4\_KINASE\_2; 1.  
CC PROSITE; PS50290; PI3\_4\_KINASE\_3; 1.  
KW Transferase; Kinase.  
FT DOMAIN 2461..2787 PI3K/PI4K.  
FT CONFLICT 1190..1190 Y -> F (IN REF. 2).  
SQ SEQUENCE 2787 AA; 321663 MW; 439B6E189E39499B CRC64;

Query Match 64.4%; Score 29; DB 1; Length 2787;  
Best Local Similarity 62.5%; Pred. No. 4.2e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLS 8  
:|||||  
Db 1996 LGSIIQLA 2003

RESULT 33  
G3P\_MYCGE STANDARD; PRT; 337 AA.  
ID G3P\_MYCGE  
AC P47543;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).  
GN GAPA OR GAP OR MG301.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).  
RN [2]  
RP SEQUENCE OF 1-81 AND 279-337 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random  
RT sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE  
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
CC -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
CC DEHYDROGENASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U39710; AAC71523.1; -  
CC EMBL; U02213; AAD12507.1; -  
CC EMBL; U02178; AAD12463.1; -  
CC HSSP; P17721; 1HDG.  
CC TIGR; MG301; -  
CC InterPro; IPR000173; -  
CC Pfam; PF00044; gpdh; 1.  
CC PRINTS; PR00078; G3PDHDRGNASE.  
CC PROSITE; PS00071; GAPDH; 1.  
KW Glycolysis; Oxidoreductase; NAD.  
FT BINDING 157..157 GLYCERALDEHYDE 3-PHOSPHATE  
FT (BY SIMILARITY).  
FT ACT\_SITE 184..184 ACTIVATES THIOL GROUP DURING CATALYSIS  
FT (BY SIMILARITY).  
SQ SEQUENCE 337 AA; 37097 MW; FA1E1966687006B CRC64;

Query Match 62.2%; Score 28; DB 1; Length 337;  
Best Local Similarity 85.7%; Pred. No. 66;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIIVQLS 8  
|||||



Db 244 GSIVELS 250

RESULT 34  
SED5\_YEAST  
ID SED5\_YEAST STANDARD; PRT; 340 AA.  
AC Q01590;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE INTEGRAL MEMBRANE PROTEIN SED5.  
GN SED5 OR YLR026C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=1400588;  
RX MEDLINE=93016264; PubMed=1400588;  
RA Hardwick K.G., Pelham H.R.B.;  
RT "SED5 encodes a 39-kD integral membrane protein required for  
RT vesicular transport between the ER and the Golgi complex.";  
RL J. Cell Biol. 119:513-521(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Obermaier B., Piravandi E., Rinke M.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE  
CC ENDOPLASMIC RETICULUM AND THE GOLGI COMPLEX. ACTS AS A TARGET  
CC ORGANELLE SOLUBLE NSF ATTACHMENT PROTEIN RECEPTOR (T-SNARE).  
CC -!- SUBUNIT: INTERACTS WITH STP1 AND SFB3.  
CC -!- SIMILARITY: BELONGS TO THE SYNTAXIN/EPIMORPHIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X66980; CAA47390.1;  
DR EMBL; 273198; CAA97549.1;  
DR PIR; A44019; A44019.  
DR PIR; S23223; S23223.  
DR SGD; S0004016; SED5.  
DR InterPro; IPR000017;  
DR Pfam; PF00804; Syntaxin; 1.  
DR PROSITE; PS00914; SYNTAXIN; 1.  
KW Coiled coil; Transport; Protein transport; Transmembrane; Golgi stack.  
FT DOMAIN 146 173 COILED COIL (POTENTIAL).  
FT TRANSMEM 320 340 POTENTIAL.  
SQ SEQUENCE 340 AA; 38807 MW; B3192B57269A0B83 CRC64;  
  
Query Match 62.2%; Score 28; DB 1; Length 340;  
Best Local Similarity 85.7%; Pred. No. 67;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 SIVQLSQ 9  
I:|||||  
Db 109 SLVQLSQ 115  
  
RESULT 35  
SCN1\_SCHPO  
ID SCN1\_SCHPO STANDARD; PRT; 387 AA.  
AC P41890;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE SCN1 PROTEIN.  
GN SCN1.

OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=7798319;  
RX MEDLINE=95096177; PubMed=7798319;  
RA Samejima I., Yanagida M.;  
RT "Bypassing anaphase by fission yeast cut9 mutation: requirement of  
RT cut9+ to initiate anaphase.";  
RL J. Cell Biol. 127:1655-1670(1994).  
CC -!- FUNCTION: INTERACTS WITH CUT9.  
CC -!- SIMILARITY: TO YEAST YMR262W.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D31845; BAA06631.1;  
DR InterPro; IPR001130;  
DR Pfam; PF01026; UPF0006; 1.  
SQ SEQUENCE 387 AA; 44421 MW; AAC154B805BC1085 CRC64;  
  
Query Match 62.2%; Score 28; DB 1; Length 387;  
Best Local Similarity 75.0%; Pred. No. 77;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 GSIVOLSQ 9  
I:|||||  
Db 294 GSIEQISQ 301  
  
RESULT 36  
COAT\_BBV  
ID COAT\_BBV STANDARD; PRT; 407 AA.  
AC P04329;  
DT 20-MAR-1987 (Rel. 04, Created)  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE COAT PROTEIN PRECURSOR.  
GN ALPHA.  
OS Black beetle virus (BBV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nodaviridae;  
OC Alphavirus.  
OX NCBI\_TaxID=12285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85014162; PubMed=6548308;  
RA Dasgupta R., Ghosh A., Dasmahapatra B., Guarino L.A., Kaesberg P.;  
RT "Primary and secondary structure of black beetle virus RNA2, the  
RT genomic messenger for BBV coat protein precursor.";  
RL Nucleic Acids Res. 12:7215-7223(1984).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY, AND SIMILARITY TO OTHER NODAVIRUSES.  
RX MEDLINE=90339486; PubMed=2116525;  
RA Kaesberg P., Dasgupta R., Sgro J.-Y., Wery J.-P., Selling B.H.;  
RA Hosur M.V., Johnson J.E.;  
RT "Structural homology among four nodaviruses as deduced by sequencing  
RT and X-ray crystallography.";  
RL J. Mol. Biol. 214:423-435(1990).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
RX MEDLINE=94118310; PubMed=8289282;  
RA Wery J.-P., Reddy V.S., Hosur M.V., Johnson J.E.;  
RT "The refined three-dimensional structure of an insect virus at 2.8-A  
RT resolution.";  
RL J. Mol. Biol. 235:565-586(1994).

CC -!- PTM: ENZYMATICALLY CLEAVED INTO COAT PROTEINS BETA AND GAMMA.  
CC HOWEVER, THE EXACT CLEAVAGE SITE HAS NOT BEEN DETERMINED.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A6.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X00956; CAA25468.1; -  
DR PIR; A04151; VCB22G.  
DR PIR; S11036; S11036.  
DR PDB; 2BBV; 31-AUG-94.  
DR MEROPS; A06.001; -  
DR InterPro; IPR000696; -  
DR Pfam; PF01829; Peptidase\_A6; 1.  
DR PRINTS; PR00863; NODAVIRPTASE.  
KW Coat protein; Hydrolase; Aspartyl protease; 3D-structure.  
SQ SEQUENCE 407 AA; 43838 MW; 3A6C6C9A98A5C26C CRC64;  
  
Query Match 62.2%; Score 28; DB 1; Length 407;  
Best Local Similarity 55.6%; Pred. No. 82;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MGSIVQLSQ 9  
|||: |||  
Db 53 MGALTRLSQ 61  
  
RESULT 37  
DCTA\_ECOLI  
ID DCTA\_ECOLI STANDARD; PRT; 428 AA.  
AC P37312;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE C4-DICARBOXYLATE TRANSPORT PROTEIN.  
GN DCTA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=94316500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
RT region from 76.0 to 81.5 minutes."  
RL Nucleic Acids Res. 22:2576-2586(1994).  
CC -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF DICARBOXYLATES SUCH AS  
CC SUCCINATE, FUMARATE, AND MALATE FROM THE PERIPLASM ACROSS THE  
CC INNER MEMBRANE (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY  
CC (SDF).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U00039; AAB18505.1; -  
DR EMBL; AE000429; AAC76553.1; -  
DR EcoGene; EG20044; dcta.  
DR InterPro; IPR001991; -

DR Pfam; PF00375; SDF; 1.  
DR PROSITE; PS00713; NA\_DICARBOXYL\_SYMPT\_1; 1.  
DR PROSITE; PS00714; NA\_DICARBOXYL\_SYMPT\_2; 1.  
KW Transport; Transmembrane; Inner membrane; Symport.  
FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 21 39 1 (POTENTIAL).  
FT DOMAIN 40 58 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 59 77 2 (POTENTIAL).  
FT DOMAIN 78 89 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 90 108 3 (POTENTIAL).  
FT DOMAIN 109 139 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 140 158 4 (POTENTIAL).  
FT DOMAIN 159 161 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 162 180 5 (POTENTIAL).  
FT DOMAIN 181 199 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 200 218 6 (POTENTIAL).  
FT DOMAIN 219 231 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 232 250 7 (POTENTIAL).  
FT DOMAIN 251 271 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 272 290 8 (POTENTIAL).  
FT DOMAIN 291 310 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 311 329 9 (POTENTIAL).  
FT DOMAIN 330 340 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 341 359 10 (POTENTIAL).  
FT DOMAIN 360 368 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 369 388 11 (POTENTIAL).  
FT DOMAIN 389 395 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 396 414 12 (POTENTIAL).  
FT DOMAIN 415 428 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 428 AA; 45436 MW; D9B32F987B62D234 CRC64;

Query Match 62.2%; Score 28; DB 1; Length 428;  
Best Local Similarity 55.6%; Pred. No. 86;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9  
|||: |||  
Db 213 VGTLVQLGQ 221

RESULT 38  
SMVA\_SALTY  
ID SMVA\_SALTY STANDARD; PRT; 496 AA.  
AC P37594;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE METHYL VIOLOGEN RESISTANCE PROTEIN SMVA.  
GN SMVA.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SL1303;  
RX MEDLINE=95011654; PubMed=7926834;  
RA Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,  
RA Ichimura S., Noda Y.;  
RT "The methyl viologen-resistance-encoding gene smva of Salmonella  
RT typhimurium."  
RL Gene 148:173-174(1994).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN  
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D26057; BAA05059.1; -  
DR StyGene; SG10384; smvA;  
KW Transmembrane; Inner membrane.  
FT TRANSMEM 5 25 POTENTIAL.  
FT TRANSMEM 44 64 POTENTIAL.  
FT TRANSMEM 73 93 POTENTIAL.  
FT TRANSMEM 96 116 POTENTIAL.  
FT TRANSMEM 135 155 POTENTIAL.  
FT TRANSMEM 158 178 POTENTIAL.  
FT TRANSMEM 192 212 POTENTIAL.  
FT TRANSMEM 220 240 POTENTIAL.  
FT TRANSMEM 260 280 POTENTIAL.  
FT TRANSMEM 299 319 POTENTIAL.  
FT TRANSMEM 327 347 POTENTIAL.  
FT TRANSMEM 357 377 POTENTIAL.  
FT TRANSMEM 391 411 POTENTIAL.  
FT TRANSMEM 470 490 POTENTIAL.  
FT SEQUENCE 496 AA; 52521 MW; 0F1B23C8FD27BCCD CRC64;  
  
Query Match 62.2%; Score 28; DB 1; Length 496;  
Best Local Similarity 60.0%; Pred. NO. 1e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 3; Gaps 0;  
  
OY 1 MGSIVQLSQS 10  
Db 437 MGEAVQLANS 446  
  
RESULT 39  
CPJ5\_MOUSE STANDARD; PRT; 501 AA.  
AC 054749;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE CYTOCHROME P450 2J5 (EC 1.14.14.1) (CYP11J5) (ARACHIDONIC ACID  
DE EPOXYGENASE).  
GN CYP2J5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS7BL/6 X CBA; TISSUE=Liver;  
RX MEDLINE=98234557; PubMed=9570962;  
RA Ma J., Ramachandran S., Fiedorek F.T. Jr., Zeldin D.C.;  
RT "Mapping of the CYP2J cytochrome P450 genes to human chromosome 1 and  
RT mouse chromosome 4";  
RL Genomics 49:152-155(1998).  
CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +  
CC -1- OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U62294; AAB87635.1; -  
DR MGD; MGI:1270149; CYP2j5.  
DR InterPro; IPR001128; -  
DR InterPro; IPR002401; -  
DR Pfam; PF00067; p450; 1.  
DR PRINTS; PR00385; P450.  
DR -----

DR PRINTS; PR00463; EP4501.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT BINDING 447 447 HEME (BY SIMILARITY).  
SQ SEQUENCE 501 AA; 57784 MW; C67F2E79DD64AF99 CRC64;  
  
Query Match 62.2%; Score 28; DB 1; Length 501;  
Best Local Similarity 60.0%; Pred. NO. 1e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
OY 1 MGSIVQLSQS 10  
Db 371 MGNIVPLNSS 380  
  
RESULT 40  
YHVO\_YEAST STANDARD; PRT; 579 AA.  
ID YHVO\_YEAST  
AC P38848;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL 66.1 KDA PROTEIN IN IMP3-SPO12 INTERGENIC REGION.  
GN YHR150W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
RA Kucaba T., Hillier L., Jler M., Johnston L., Langston Y., Mouser L.,  
RA Latreille P., Louis E.J., Macris C., Mardis E., Menezes S., Vaughan K.,  
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Wilson R.,  
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
RA Vaudin M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
RT VIII";  
RL Science 265:2077-2082(1994).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U10397; AAB68980.1; -  
DR PIR; S46757; S46757.  
DR SGD; S0001193; YHR150W.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 237 253 POTENTIAL.  
FT TRANSMEM 397 421  
SQ SEQUENCE 579 AA; 66148 MW; 2A6A73999C54C19E CRC64;  
  
Query Match 62.2%; Score 28; DB 1; Length 579;  
Best Local Similarity 60.0%; Pred. NO. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
OY 1 MGSIVQLSQS 10  
Db 219 MGSIFELQDS 228  
  
RESULT 41  
YEHQ\_ECOLI

```
ID YEHO_ECOLI STANDARD; PRT; 614 AA.
AC P33353;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 67.7 KDA PROTEIN IN MOLR-BGLX INTERGENIC REGION.
GN YEHO.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Raley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00007; AAA60485.1; ALT_INIT.
DR EMBL; AE000301; AAC75183.1; ALT_INIT.
DR EcoGene; EG12003; yehQ.
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 67730 MW; 8056294BAE3CAE6E CRC64;

Query Match 62.2%; Score 28; DB 1; Length 614;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
   |||: ||:|
Db 365 GSILHLSRS 373

RESULT 42
KDGL_DROME STANDARD; PRT; 791 AA.
AC Q01583;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIACYLGLYCEROL KINASE (EC 2.7.1.107) (DIGLYCERIDE KINASE 1) (DGK 1)
DE (DAG KINASE 1).
GN DGK OR DGK1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=92335231; PubMed=1321433;
RA Masai I., Hosoya T., Kojima S., Hotta Y.;
RT "Molecular cloning of a Drosophila diacylglycerol kinase gene that is
```

```
RT expressed in the nervous system and muscle.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6030-6034(1992).
RN [2]
RN SEQUENCE OF 279-791 FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=93143713; PubMed=8380995;
RA Harden N., Yap S.F., Chiam M.-A., Lim L.;
RT "A Drosophila gene encoding a protein with similarity to
RL diacylglycerol kinase is expressed in specific neurons.";
RL Biochem. J. 289:439-444(1993).
CC -!- FUNCTION: UPON CELL STIMULATION CONVERTS THE SECOND MESSENGER
CC DIACYLGLYCEROL INTO PHOSPHATIDATE, INITIATING THE RESYNTHESIS
CC OF PHOSPHATIDYLINOSITOLS AND ATTENUATING PROTEIN KINASE C
CC ACTIVITY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL = ADP +
CC 1,2-DIACYLGLYCEROL 3-PHOSPHATE.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE NERVOUS SYSTEM AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: DGK IS TRANSCRIBED IN THE EMBRYONIC, PUPAL
CC AND ADULT STAGES, WITH LITTLE EXPRESSION DURING THE LARVAL
CC STAGES. EXPRESSION IN LATE EMBRYOS IS SPECIFIC TO THE CENTRAL
CC NERVOUS SYSTEM AND HEAD.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D11120; BAA01894.1; -.
DR EMBL; X67335; CAA47750.1; -.
DR PIR; S25099; S25099.
DR PIR; S28229; S28229.
DR PIR; A46140; A46140.
DR FlyBase; FBgn0004568; Dgk.
DR InterPro; IPR000756; -.
DR InterPro; IPR001206; -.
DR Pfam; PF00609; DAGKa; 1.
DR Pfam; PF00781; DAGKc; 1.
DR Transferase; Kinase.
KW DOMAIN 344 353 THR-RICH.
FT DOMAIN 401 405 POLY-ALA.
FT DOMAIN 430 444 GLN-RICH.
FT DOMAIN 534 539 POLY-GLN.
FT DOMAIN 130 256 CATALYTIC-A (POTENTIAL).
FT DOMAIN 564 755 CATALYTIC-B (POTENTIAL).
FT CONFLICT 374 374 T -> M (IN REF. 2).
FT CONFLICT 444 444 Q -> QQQQ (IN REF. 2).
SQ SEQUENCE 791 AA; 87267 MW; 7D50D8ED01496679 CRC64;

Query Match 62.2%; Score 28; DB 1; Length 791;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
   |:::|
Db 279 MQKVIELSQS 288

RESULT 43
PERT_BORPE STANDARD; PRT; 910 AA.
ID PERT_BORPE
AC P14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PERTACTIN PRECURSOR (OUTER MEMBRANE PROTEIN P.69) (P.93).
GN PRN OR OMP69A.
OS Bordetella pertussis.
```



OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN-ISOLATE CN2992;  
RX MEDLINE=89264462; PubMed=2542937;  
RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,  
RT Novotny P., Morrissey P., Fairweather N.F.,  
RT "Molecular cloning and characterization of protective outer membrane  
RT protein P.69 from Bordetella pertussis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).  
RN [2]  
RN REVISIONS TO 264 AND 332.  
RX MEDLINE=92407514; PubMed=1527510;  
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.,  
RT "Cloning, nucleotide sequence and heterologous expression of the  
RT protective outer-membrane protein P.68 pertactin from Bordetella  
RT bronchiseptica.";  
RL J. Gen. Microbiol. 138:1697-1705(1992).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=96196517; PubMed=8609998;  
RA Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.,  
RT "Structure of Bordetella pertussis virulence factor P.69 pertactin.";  
RL Nature 381:90-92(1996).  
CC -!- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS  
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN  
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.  
CC -!- SUBUNIT: MONOMER.  
CC -!- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.  
CC -!- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)  
CC CONCENTRATION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J04560; AAA22980.1; ALT\_SEQ.  
DR PIR: A32560; A32560.  
KW Outer membrane; Signal; Virulence; Repeat.  
FT SIGNAL 1 34  
FT CHAIN 35 910 P.93.  
FT CHAIN 35 911 PERTACTIN (P.69).  
FT PROPEP 712 910 POTENTIAL.  
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN  
FT ADHESION TO VARIOUS EUKARYOTIC CELL  
FT LINES).  
FT DOMAIN 266 290 5 X 5 AA TANDEM REPEATS OF G-G-A-V-P.  
FT REPEAT 266 270 1.  
FT REPEAT 271 275 2.  
FT REPEAT 276 280 3.  
FT REPEAT 281 285 4 (APPROXIMATE).  
FT REPEAT 286 290 5 (APPROXIMATE).  
FT DOMAIN 579 593 5 X 3 AA TANDEM REPEATS OF P-Q-P.  
SQ SEQUENCE 910 AA; 93452 MW; A169871E20A2E7DB CRC64;

Query Match 62.2%; Score 28; DB 1; Length 910;  
Best Local Similarity 66.7%; Pred. No. 2e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GSIVQLSQS 10  
Db 302 GSSVELAQ 310  
RESULT 44

TYCA\_BACBR STANDARD; PRT; 1088 AA.  
AC P09095; O30407;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE TYROCIDINE SYNTHETASE I [INCLUDES: ATP-DEPENDENT D-PHENYLALANINE  
DE ADENYLASE (D-PHEA) (D-PHENYLALANINE ACTIVASE); PHENYLALANINE RACEMASE  
DE [ATP-HYDROLYZING] (EC 5.1.1.11)].  
GN TYCA.  
OS Bacillus brevis.  
OS Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Brevibacillus.  
OC NCBI\_TaxID=1393;  
OX [1]  
RX SEQUENCE FROM N.A.  
RC STRAIN=ATCC 8185;  
RX MEDLINE=89098354; PubMed=3267240;  
RA Weckermann R., Fuerbass R., Marahiel M.A.,  
RT "Complete nucleotide sequence of the tyca gene coding the tyrocidine  
RT synthetase 1 from Bacillus brevis.";  
RL Nucleic Acids Res. 16:11841-11841(1988).  
RN [2]  
RX SEQUENCE FROM N.A.  
RC STRAIN=ATCC 8185;  
RX MEDLINE=98012987; PubMed=9352938;  
RA Mootz H.D., Marahiel M.A.,  
RT "The tyrocidine biosynthesis operon of Bacillus brevis: complete  
RT nucleotide sequence and biochemical characterization of functional  
RT internal adenylation domains.";  
RL J. Bacteriol. 179:6843-6850(1997).  
RN [3]  
RX SEQUENCE OF 1-62 FROM N.A.  
RA Marahiel M.A., Zuber P., Czekay G., Losick R.,  
RT "Identification of the promoter for a peptide antibiotic biosynthesis  
RT gene from Bacillus brevis and its regulation in Bacillus subtilis.";  
CC -!- FUNCTION: IN THE FIRST STEP OF PEPTIDE SYNTHESIS THIS ENZYME  
CC ACTIVATES PHENYLALANINE AND RACEMIZES IT TO THE D-ISOMER.  
CC -!- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE -> AMP + PYROPHOSPHATE  
CC + D-PHENYLALANINE.  
CC -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE  
CC (POTENTIAL).  
CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE  
CC ANTIBIOTIC TYROCIDINE.  
CC -!- SUBUNIT: LARGE MULTIZYME COMPLEX OF TYCA, TYCB AND TYCC.  
CC -!- DOMAIN: ONE-MODULE-BEARING PEPTIDE SYNTHASE WITH A C-TERMINAL  
CC EPIMERIZATION DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO  
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS  
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION  
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND  
CC N METHYLATION (OPTIONAL).  
CC -!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC  
CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-  
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR  
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE  
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.  
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT  
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE. 56% IDENTITY TO  
CC GRAMICIDIN S SYNTHETASE I (GRSA), WHICH CATALYZES THE SAME  
CC REACTION IN GRAMICIDIN S SYNTHESIS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X13237; CAA31623.1; -.  
DR EMBL: AF004835; AAC45928.1; -.

```
DR EMBL; M16442; AAA22876.1; ALT_FRAME.
DR PIR; S02039; YGBSTB.
DR PIR; A26878; A26878.
DR HSSP; P14687; IAMU.
DR InterPro; IPR000255; -.
DR InterPro; IPR000873; -.
DR InterPro; IPR001242; -.
DR Pfam; PF00501; AMP-binding; 1.
DR Pfam; PF00668; DUF4; 1.
DR Pfam; PF00550; pp-binding; 1.
DR PRINTS; PRO0154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR PROSITE; PS00455; AMP_BINDING; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
KW Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme.
FT DOMAIN 533 599 ACYL CARRIER (ACP).
FT BINDING 563 563 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT CONFLICT 197 242 NLOSFFQNSFGVTEQDRIGLFASMSFDASVWEMFMALLSGA
FT CONFLICT 359 360 SLVIL -> ICNPFKIRLASPSKTSKSGFLPACRSTHPFGK
FT CONFLICT 665 665 CSWLCLAPRVHP (IN REF. 1).
FT CONFLICT 737 737 GS -> AD (IN REF. 1).
FT CONFLICT 756 790 L -> V (IN REF. 1).
FT CONFLICT 737 737 L -> F (IN REF. 1).
FT CONFLICT 756 790 EDLATGYAALAGQAISLPKTDTSFQSWSQLQY ->
FT CONFLICT 756 790 KIWPQDTRRHLLQKRSVCPKPKRILFKAGHNGCKNN (IN
FT CONFLICT 876 894 REF. 1).
FT CONFLICT 876 894 QIVIHLEGHGREDIIEQAN -> KSSFIWRGTGARTSSNRQ
FT CONFLICT 876 894 T (IN REF. 1).
SQ SEQUENCE 1088 AA; 122672 MW; 6AC4804199572027 CRC64;

Query Match 62.2%; Score 28; DB 1; Length 1088;
Best Local Similarity 87.5%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 SIVQLSQS 10
|| |||||
Db 1034 SIQQLSQS 1041

RESULT 45
MNN4_YEAST STANDARD; PRT; 1178 AA.
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MNN4 PROTEIN.
GN MNN4 OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97175967; PubMed=9023541;
RA Odani T., Shlitta Y.-I., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MNN4 gene required for phosphorylation
RT of N-linked oligosaccharides in Saccharomyces cerevisiae.";
RL Glycobiology 6:805-810(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Maia E Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
CC MANNOSYLPHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
CC MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
CC PORTIONS OF N-LINKED OLIGOSACCHARIDES.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: TO YEAST YJR061W.
```

```
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D83006; BA11676.1; -.
DR EMBL; Z28201; CAA82046.1; -.
DR EMBL; Z28200; CAA82044.1; -.
DR PIR; S38037; S38037.
DR PIR; S38038; S38038.
DR YEPD; 6530; -.
DR SGD; S0001683; MNN4.
KW Transmembrane; Signal-anchor; Repeat.
FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 49 1178 (POTENTIAL).
FT DOMAIN 1032 1174 LUMENAL (POTENTIAL).
FT DOMAIN 1042 1174 ARG/GLU/LYS-RICH (HIGHLY CHARGED).
FT REPEAT 1042 1049 17 X 8 AA TANDEM REPEATS OF K-K-K-E-E-
FT REPEAT 1049 1049 E-E.
FT REPEAT 1050 1057 1.
FT REPEAT 1058 1065 2.
FT REPEAT 1066 1073 3.
FT REPEAT 1074 1081 4.
FT REPEAT 1082 1089 5.
FT REPEAT 1090 1097 6.
FT REPEAT 1098 1105 7 (APPROXIMATE).
FT REPEAT 1106 1113 8.
FT REPEAT 1114 1121 9 (APPROXIMATE).
FT REPEAT 1122 1129 10 (APPROXIMATE).
FT REPEAT 1130 1137 11 (APPROXIMATE).
FT REPEAT 1138 1144 12.
FT REPEAT 1145 1152 13 (APPROXIMATE).
FT REPEAT 1153 1160 14 (APPROXIMATE).
FT REPEAT 1161 1168 15 (APPROXIMATE).
FT REPEAT 1169 1174 16 (APPROXIMATE).
FT DOMAIN 37 40 POLY-ILE.
FT DOMAIN 1042 1045 POLY-LYS.
FT DOMAIN 1046 1049 POLY-GLU.
FT DOMAIN 1050 1053 POLY-LYS.
FT DOMAIN 1054 1057 POLY-GLU.
FT DOMAIN 1058 1061 POLY-LYS.
FT DOMAIN 1062 1065 POLY-GLU.
FT DOMAIN 1066 1069 POLY-LYS.
FT DOMAIN 1070 1073 POLY-GLU.
FT DOMAIN 1074 1077 POLY-LYS.
FT DOMAIN 1078 1081 POLY-GLU.
FT DOMAIN 1082 1085 POLY-LYS.
FT DOMAIN 1086 1089 POLY-GLU.
FT DOMAIN 1094 1097 POLY-GLU.
FT DOMAIN 1098 1101 POLY-LYS.
FT DOMAIN 1102 1105 POLY-GLU.
FT DOMAIN 1134 1137 POLY-GLU.
FT DOMAIN 1157 1160 POLY-GLU.
FT DOMAIN 1165 1168 POLY-GLU.
SQ SEQUENCE 1178 AA; 139380 MW; BC05DAEOAEFCB282 CRC64;

Query Match 62.2%; Score 28; DB 1; Length 1178;
Best Local Similarity 50.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
:| | :| |
Db 931 VGKIIDLSES 940
```

- Search completed: June 28, 2001, 11:57:31  
Job time: 272 sec

---